Technical Report 1074

Combat Vehicle Training with Thermal Imagery

Jean L. Dyer
U.S. Army Research Institute

Amy J. Westergren and George W. Shorter
Auburn University
Consortium Research Fellows Program

Larry D. Brown
Bradley Instructional Company, 29th Infantry Regiment

November 1997

United States Army Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Technical review by

Larry L. Meliza, ARI SSRU
Margaret S. Salter, ARI IFRU

NOTICES

~~**DISTRIBUTION**: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.~~

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1997, November	2. REPORT TYPE Final	3. DATES COVERED (from... to) June 1996-June 1997
4. TITLE AND SUBTITLE Combat Vehicle Training with Thermal Imagery		5a. CONTRACT OR GRANT NUMBER 5b. PROGRAM ELEMENT NUMBER 062875A
6. AUTHOR(S) Jean L. Dyer (ARI), Amy J. Westergren (Auburn University), George W. Shorter (Auburn University), and Larry D. Brown (29th Infantry)		5c. PROJECT NUMBER A790 5d. TASK NUMBER 2223 5e. WORK UNIT NUMBER H01
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences Infantry Forces Research Unit P.O. Box 52086 Fort Benning, GA 31995-2086		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Technical Report 1074

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

Three training effectiveness experiments were conducted on a computer-based, thermal combat vehicle identification program developed under the auspices of the Product Manager for Forward Looking Infrared. The program included real thermal imagery of vehicles. The first experiment compared part-task training schedules. The second compared a self-paced trial procedure, which incorporated side-by-side visual corrective feedback, to a timed trial procedure, which gave only knowledge of results feedback. The third examined the effects of training at near versus far ranges.

Thermal training substantially improved scores on both thermal and visible images of vehicles that had been trained, but not for other vehicles. Long vehicle sets led to learning plateaus; shorter and multiple sets worked better. Learning was more efficient and transfer was enhanced when soldiers responded at their own pace and received corrective visual feedback. Soldiers learned to discriminate vehicles at far ranges, although it took them twice as long as soldiers who trained on near imagery. Even with extensive training, some vehicle confusions persisted, indicating great similarity in thermal signatures for some vehicles. The findings were applied to the program. Efforts are continuing to refine it and to field it throughout the Army.

15. SUBJECT TERMS

Training	Recognition	Combat Vehicle Identification	Identification
Thermal	Computer-based training	Target Acquisition	FLIR

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	198	Jean L. Dyer (706) 545-1278

Technical Report 1074

Combat Vehicle Training with Thermal Imagery

Jean L. Dyer
U.S. Army Research Institute

Amy J. Westergren and George W. Shorter
Auburn University
Consortium Research Fellows Program

Larry D. Brown
Bradley Instructional Company, 29th Infantry Regiment

Infantry Forces Research Unit
Scott E. Graham, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

November 1997

Army Project Number
20262785A790

Education and Training Technology

FOREWORD

Technological advances in soldier equipment usually require new training materials to enable soldiers to employ the equipment to its utmost. This was especially the case when forward looking infrared (FLIR) weapon sights (aka thermal) were introduced in the 1970s. This technology provided soldiers a different way of viewing the battlefield at night as well as during the day. Thermal signatures emitted by vehicles were not the same as their visible counterparts. Target acquisition required new skills. Unfortunately, training materials did not accompany this technology. And with the proliferation of thermal sights in the Army, the need for training materials has increased.

In 1995 the Product Manager for Second Generation FLIR (PM-FLIR) sponsored a training effort to develop a target acquisition program for 2nd-generation FLIR sights. The primary training objective was to provide soldiers with combat vehicle identification skills, but training them to detect vehicles was also an objective. The critical and unique aspect of the program was that actual thermal imagery was to be used. Soldiers would see the "real thing," not computer-generated imagery that simulated, but failed to replicate, the real world. The program applies to 1st-generation FLIR sights as well. The Army's Night Vision and Electronic Sensors Directorate (NVESD) developed the program to include obtaining day and night imagery of a large number of combat vehicles and developing the computer-based training program.

The Infantry Forces Research Unit (IFRU) of the U.S. Army Research Institute for the Behavioral and Social Sciences, in coordination with PM-FLIR and NVESD, conducted training effectiveness experiments on a prototype version of the program. This technical report describes these experiments and how the findings were used to improve the instructional and training design of the program.

The program is under continuing development, expansion, and refinement as additional resources become available. Beta sites in Infantry and armor units were established in 1997 to obtain additional soldier input on prototype versions of the program. The Infantry School is currently reviewing the program, and the Dismounted Battlespace Battle Lab continues to use the program in its Night Fighting Experimentation Facility. In 1997, formal delivery of the program was made to PM-FLIR by the NVESD.

ZITA M. SIMUTIS
Technical Director

ACKNOWLEDGMENT

The research was made possible with the support of the Product Manager-Second Generation Forward Looking Infrared (PM-FLIR) and colleagues at the Army's Night Vision and Electronic Sensors Directorate (NVESD). LTG J. Mackin and Greg Mann, PM-FLIR, provided the resources for the experiments. Dr. Barbara O'Kane, Geoffrey Sawyer, and John Palmer at NVESD backed the entire effort and worked closely with us in incorporating the lessons learned from the experiments in the delivery package to PM-FLIR. We also appreciate the assistance provided by the professional staff at E-OIR Measurements, Inc. who addressed emergency problems we encountered with the training software.

Finally, we wish to thank the many Infantry soldiers and leaders who participated in the experiments. Their dedication, high level of motivation, and sincere interest in learning to identify the vehicles made the experiments a success.

COMBAT VEHICLE TRAINING WITH THERMAL IMAGERY

EXECUTIVE SUMMARY

Research Requirement:

Effective employment of thermal weapon sights requires special training. The thermal signatures generated by vehicles on the battlefield are not the same as their visible counterparts; nor is the image of the battlefield itself. Training combat vehicle identification skills with thermal imagery can not only help soldiers understand the dynamics of this technology as it affects battlefield operations, but such skills can also help reduce fratricide. This report describes three training effectiveness experiments on a prototype, computer-based, combat vehicle identification program, called CVI*Plus*, developed by the Night Vision and Electronic Sensors Directorate (NVESD) and sponsored by the Product Manager for Forward Looking Infrared (PM-FLIR). As no program existed prior to this time, the experiments were conducted to determine whether the program's modules should be changed, how they should be modified to improve soldier performance, and what additional capabilities were needed.

Procedure:

Three experiments were conducted with the prototype version CVI*Plus* program. The first compared part-task training schedules. The second compared a self-paced training procedure with side-by-side visual corrective feedback to a fixed-paced training procedure with the usual knowledge of result feedback (correct or incorrect). The third experiment examined the effects of training at near versus far ranges. The same experimental paradigm was used throughout. Pretests using thermal and visible images were given, followed by group instruction on thermal cues for the vehicles being trained. Then soldiers were given a series of exercises where they were required to train to a predetermined standard; the exercise format was tailored to address the objectives of each experiment. Following training, soldiers were tested for transfer. Finally, soldiers were posttested on the thermal and visible images. The pretests and posttests were identical in all experiments. They contained vehicles in the training exercises as well as vehicles that were not trained. Participants were from active Army units, the Bradley Leader's course at the Infantry School, and a National Guard Unit. In each experiment, soldiers were randomly assigned to the experimental conditions.

Findings:

Thermal training improved scores on both the thermal and visible images of the vehicles included in the training, but not the score for other, non-trained vehicles. Faster response times on the tests and training exercises also indicated greater skill. Initial individual differences remained after training, despite the substantial improvement in thermal skills. Long vehicle sets led to learning plateaus and learner frustration; shorter and multiple sets worked better. Learning

was more efficient and transfer was enhanced when soldiers were allowed to respond to the images at their own pace and received corrective visual feedback. Soldiers did learn to discriminate vehicles at far ranges, although it took them twice as long as soldiers who trained on near imagery. Not all transfer conditions were of equal difficulty: day thermal imagery was harder than night thermal imagery; front and rear aspects were hardest, followed by the oblique views, with flanks the easiest. Even with extensive training, some vehicle confusions remained for certain aspects and at certain ranges. The results also showed that even after learning to discriminate vehicles within vehicle sets to a high criterion, cross-set confusions occurred when all vehicles were pooled. Use of actual thermal imagery (in conjunction with multi-media, computer-based, practical exercises that adapt to each soldier's learning rate and can continually challenge soldier expertise) is clearly the way to make the difficult task of vehicle discrimination more interesting, effective, and efficient.

Utilization of Findings:

Findings were used to revise the CVI*Plus* program. The major impacts were: the creation of vehicle sets that can be programmed in terms of difficulty and vehicle similarity; end-of-exercise feedback on vehicle confusions; the application of pass criteria before allowing individuals to progress to the next set of vehicles; addition of an instructional module on thermal cues including comparisons between visual and thermal signatures; and accessibility to the library of all vehicle images at any point in the program. As the findings showed substantial value for both the fixed-paced and self-paced training formats and for visual corrective feedback, these features were retained with only minimal enhancements. An instructor control matrix, to enable instructors to design training sequences to meet their mission needs, was added. In addition, the report provides guidance to training developers and instructors regarding development of vehicle identification training programs, and delineates practical and theoretical questions that emerged from the research.

COMBAT VEHICLE TRAINING WITH THERMAL IMAGERY

CONTENTS

LIST OF TABLES

Page

LIST OF TABLES (continued)

Page

LIST OF TABLES (continued)

Page

LIST OF FIGURES

Page

LIST OF FIGURES (continued)

Combat Vehicle Training with Thermal Imagery

Introduction

Background

When the first thermal night sights for anti-armor weapon systems were fielded in the 1970s under the Army's Manportable Common Thermal Night Sight program, no materials were developed to train gunners on target acquisition skills with this new technology. Gunnery training devices were developed, but the training of target acquisition skills, such as target detection and vehicle recognition and identification, was not addressed. Also lacking were materials on the characteristics of this forward looking infrared (FLIR) technology; fundamental concepts which would enable soldiers to better interpret and understand the dynamic imagery seen in thermal sights under diverse atmospheric and combat conditions. Training materials were produced by the Night Vision Laboratory (Orentas, Zegel, & Gonzalez, 1991; Palmer, D'Agostino, & Lillie, 1982) and the Army Research Institute (Rollier, Champion, Roberson, & Graber, 1988; Smith, Shope, Heuckeroth, Warnick & Essig, 1987) to address these problems. However, they were interim and partial solutions. Consequently, major training deficiencies have continued to the present, to include an absence of training materials for the most recent, second-generation sights.

The research presented in this report is based on an initial version of a thermal training program, called CVI*Plus*, which uses actual thermal imagery. Three experiments were conducted to identify effective means of training combat vehicle identification (CVI) skills using CVI*Plus* and to determine the extent of skill transfer to imagery of the same vehicles under different conditions. Development of the training program was sponsored by the Army's Product Manager for Forward Looking Infrared (PM-FLIR). It was executed by the Army's Night Vision and Electronic Sensors Directorate (NVESD). Simultaneously with program development, the Army Research Institute's Infantry Forces Research Unit conducted the training effectiveness research reported here. The long-term goal of the research and development program was to develop, evaluate, and field a generic training program for all the Army's thermal sights.

Structure of the Report

The introductory section of the report provides background on thermal imagery, describes the thermal training program, and summarizes the intent of each experiment. The research paradigm common to all experiments is also presented.

Each of the three experiments is described in a separate section of the report. They are reported in the order in which they were conducted.

The findings are summarized in the last section. Recommendations regarding training program design and structure are presented. The program, as delivered to PM-FLIR in 1997, is described. Questions to be addressed in future research are identified.

Thermal Sights and Imagery

The TOW and Dragon thermal night sights and later generations of sights such as those on the Abrams tank, the Bradley Fighting Vehicle (BFV), and the Apache helicopter are also known as infrared night sights, thermal sights, and FLIRs. They all detect infrared (IR) energy and convert it to visible light. The fundamental principles underlying this capability are that:

> Heated objects tend to emit infrared energy through molecular agitation. This infrared energy . . . is called radiation or thermal radiance. The magnitude of the thermal radiance emitted by an object is determined by several factors. Included in these factors are the object's temperature, its surface reflectivity, and its structural properties. Natural infrared energy is produced when objects absorb and radiate solar energy and convective heat energy from warm air currents. Man-made energy, especially in vehicles, results from fuel combustion heat generated by engines and frictional heat generated by moving parts. These last two sources cause operational target vehicles to emit large amounts of IR energy. These various thermal sources acting upon the various objects in a real-world setting produce the radiant IR energy that makes up a thermal scene. (Palmer et al., 1982, p. 1-2)

The term FLIR refers to a "general class of electro-optical devices designed to 'see' the infrared energy" (Palmer et al., 1983, p. 2-3). Throughout this report, the terms FLIR, IR, and thermal are used interchangeably.

Detecting, recognizing, and identifying vehicles with FLIR imagery is particularly difficult. Thermal images differ from photopic, that is, daytime visible, images. Expertise in acquiring vehicles during the day does not guarantee expertise with thermal sights, as critical hot spots are not visible to the unaided eye. Furthermore, FLIR imagery is dynamic, varying with the atmospheric condition, the temperature, the operating time of the vehicle, and the contrast and brightness settings used on the sight. No two FLIR images are alike. Day thermal imagery differs from night imagery. FLIR cues become less distinct as distance to the vehicle increases. FLIR cues vary with the vehicle's aspect. These factors make thermal training a challenge.

Scope of a Thermal Training Program

A comprehensive training program with thermal imagery could be quite extensive and not restricted to just discriminating or naming vehicles. It would describe how FLIR technology functions, and explain how this applies to what soldiers are likely to see on the battlefield, what might be called "thermal basics." These concepts would be applied to vehicles, terrain, and other man-made and natural objects, enabling soldiers to interpret temperature differences in the total scene as viewed through their sights. The program would include training on searching for and detecting targets, and distinguishing real from false targets under a variety of conditions.

It would have information on the capabilities and characteristics of vehicles, what might be called "vehicle basics." "Vehicle basics" would be valuable for new soldiers and a refresher for more experienced soldiers. Following that would be "vehicle thermal signature basics." This part of the program would describe the heat plant of vehicles, for different types of vehicles as well as

individual vehicles. Also included would be explanations of how the heat plant affects the vehicle's thermal signature during day and night. Training on use of system sights (polarity, brightness, contrast, focus) to aid in vehicle detection, recognition, and identification would be included. Modules instructing and training soldiers to recognize the thermal cues unique to vehicles at different ranges, aspects, times of the day and night, and atmospheric conditions would be part of the program.

And, of course, exercises to train soldiers to name and discriminate vehicles under a variety of conditions, and to test their skill would be included. These exercises would provide soldiers with feedback on their progress, and adapt to their rate of progress. Sufficient imagery would be available so the test imagery would differ from the imagery used in training.

The prototype version of CVI*Plus* used in the research did not have all these features. The long-term research and development goal was, however, to include them. The version we used had modules that focused on training and testing soldiers to discriminate and name combat vehicles.

Target Acquisition Terms

There is a lack of consistency in how terms associated with target acquisition are defined. The JCS Pub.1 (Joint Chiefs of Staff, 1987) definitions are as follows:

> Detection: "the discovery by any means of the presence of a person, object or phenomenon of potential military significance" (p. 113).
>
> Recognition: "in ground combat operations, the determination that an object is similar within a category of something already known; e.g., tank, truck, man." (p. 304).
>
> Identification: "in ground combat operations, discrimination between recognizable objects as being friendly or enemy, or the name that belongs to the object as a member of a class" (p. 177).

The *Bradley Fighting Vehicle Gunnery* field manual, FM 23-4 (Department of the Army [DA], 1996), use of "identification" is consistent with the JCS Pub. 1. It contains a task entitled "identify combat vehicles" where soldiers are required to identify vehicles by their nomenclature.

However, not all terms in Army training publications are consistent with the JCS definitions. Many Army training materials use the term "recognition" in the same way that the JCS Pub.1 defines "identification," that is, to name the object or to determine if a vehicle is enemy or friendly. A previous edition of the *Soldier's Manual of Common Tasks* (DA, 1987b) included the task "recognize friendly and threat armored vehicles and aircraft." The criterion called for soldiers to "recognize" 10 out of 10 vehicles or aircraft as friendly or threat. The Army's graphic training aids (GTAs) for vehicles and aircraft, dating from the 1970s (DA, 1970, 1977, 1979, 1984b, 1985, 1987a) are labeled armored or combat vehicle recognition or visual aircraft recognition, even though they are intended to train soldiers and aviators to name vehicles and aircraft. This is also the case with the *Aviator's Recognition Manual* (DA, 1984a), that contains pictures of aircraft and

combat vehicles, and with the *Visual Aircraft Recognition* CD-ROM produced by the U.S. Army Air Defense Artillery School (ADA, 1993). Historically, Gibson (1947), in his World War II training experiments, used recognition to mean training soldiers to name specific aircraft.

Finally, there are engineering definitions associated with the design of electro-optic sensor performance, often based on the "Johnson criteria." These definitions have also influenced how the military defines target acquisition. With the Johnson criteria, "recognition" corresponds closely to that in the JCS Pub. 1; the meaning of "identification" is more ambiguous. The Johnson criteria for sensors (number of line pairs or resolvable cycles) emerged from laboratory experiments designed to "determine the resolution required of a system to perform certain target interpretation processes" (Donohue, 1991, p. 2). Biberman (as cited in Donohue) defined target acquisition terms as follows: detection - an object is present; recognition - the class to which the object belongs may be discerned (house, truck, man), and identification - the target can be described to the limit of the observer's knowledge (pickup truck, policeman). Later, the term "classification" was added and defined as "the visual act corresponding to perception of the general class of military targets, e.g., tracked versus wheeled vehicles" (Donohue, 1991, p. 8). These engineering definitions have been applied by the Army in operational tests of sensors and used to evaluate soldier performance with these sensors. For example, in the Limited User Test of the Improved Target Acquisition System (Test and Evaluation Command, 1995), detection was defined as discriminating an object of military interest from its background. Classification was defined as discriminating a target by general type (tracked versus wheeled vehicle). Recognition was defined as distinguishing between two targets of similar type (two types of tracked vehicles such as armored personnel carriers and tanks).

In this report, the term "identification" is used as defined in the JCS Pub. 1. It refers to naming the specific vehicle displayed, that is, whether it is an M1 tank, M2 BFV, or LAV-25.

The CVIPlus Training Program

Program Research and Development

Development of CVI*Plus* was funded in FY96 by the PM-FLIR for the second-generation of FLIR sights in the BFV and the Abrams tank. The program is a Windows-based computer program using digitized FLIR imagery of vehicles and scenes. The training software uses Icon Author™. An early version of the program was examined. Since the same program was used in all experiments, it is described in full in this Introduction section.

CVIPlus Imagery

The digitized FLIR imagery is central to the CVI*Plus* program. In fact, a major part of program development was the collection of this "real" imagery for a substantial number of vehicles. Simply stated, no set of images appropriate for training soldiers on the thermal signatures of combat vehicles existed prior to this collection effort. For the prototype program, the FLIR imagery was collected with an 8-12 micron calibrated, commercial thermal imager (Agema T1000). The imager was positioned 400 m from vehicles for ground-to-ground

collection, and collected in the 20° wide field of view (FOV) and the 5° narrow FOV. Vehicles were exercised for 20 min prior to imagery collection. The vehicles were positioned in eight stationary orientations with respect to the sensor; engines were idling. The eight orientations were 0, 30, 90, 150, 180, 210, 270, and 330 degrees, corresponding to the front, left front oblique, left flank, left rear oblique, rear, right rear oblique, right flank, and right front oblique. Both day and night images were collected in April and August in the U.S. temperate climate. Each 12-bit digital frame was created by averaging 16 frames. Frame averaging was used to effectively improve the imager's sensitivity to more closely match that of tactical system sights.

The raw imagery was processed in several steps to prepare it for the CVI*Plus* program. First, it was cropped to fit the window size in the program and to create a suitable target-to-background ratio. The images were then scaled to simulate views at four ranges or distances by using a sensor convolution kernal employing the narrow FOV for the close ranges and the wide FOV for the far ranges (i.e., applying the appropriate sensor spatial degradation for each range). Generic ranges (labeled Close-up, Ranges 1, 2, and 3) were used to keep the program unclassified and because no particular thermal sight was simulated. The 12-bit images were displayed in an 8-bit mode by a process that ensured the hot spot features were visible. In the field, the soldier can adjust the sight's brightness and contrast controls to bring out the hot-spot features. But in the CVI*Plus* program only one setting was available. Therefore this setting had to show features clearly to represent field capability. The images were shades of green.

The imagery in CVI*Plus* differs in two major respects from that in many gunnery trainers with night capability, such as the conduct-of-fire-trainer for the Abrams tank and the BFV. First, the imagery is actual FLIR imagery, rather than simulated, computer-generated images of vehicles. This distinction is critical, as fratricide and other combat incidents indicate the need to train vehicle identification with realistic imagery and to portray the dynamics of FLIR imagery. Second, both day and night imagery of vehicles is available. In many gunnery training devices, only the nighttime is simulated.

The imagery within CVI*Plus* also differs from current and previous vehicle and aircraft recognition/identification training programs using daytime or photopic images. A major distinguishing feature of CVI*Plus* is the systematic approach to presenting images. All vehicles are presented at four ranges, at eight aspects, at night, at day, and in white-hot and black-hot in order to show vehicle cues under a variety of conditions. With white-hot imagery, hot areas appear light or white; in black-hot, they appear dark or black. During imagery collection, great care was taken to minimize incidental cues in the background. This is an important feature, as soldiers will use incidental cues, rather than the distinguishing features of the target itself, to identify vehicles or aircraft (Gibson, 1947; Warnick & Smith, 1989). Cost limited the number of vehicles in the program. However, the projected number of vehicles in the final version, 35, approximates that in the Army's most commonly used training aid, GTA 17-2-3 (DA, 1987a) and the number in the performance test specified in the BFV FM 23-1 (DA, 1996).

By comparison, current training aids and reference materials for training daytime target identification may present targets at different ranges or at different aspects, but typically not both. Depending on the media, incidental cues may be present. Current training aids and

reference materials use line drawings (DA, 1987a, Foss, 1992, Jane's Information Group, 1997), photographs of scaled models (Hansen, 1970; Warnick & Smith, 1989), scaled models, photographs (Cullen & Foss, 1996; DA, 1984b; Foss, 1992; Foss, 1995; Foss & Gander, 1996; Jane's Information Group, 1997), or computer-generated three-dimensional models (ADA, 1993). Line drawings, pictures of scaled models, and computer-generated three-dimensional models are an inexpensive means of providing images of a large number of targets and of controlling the background. Photographs, although more realistic but often taken under fortuitous circumstances, frequently present incidental cues. The GTA 17-2-3 (DA, 1987a) shows line drawings of vehicles at one range and few aspects, typically a front, flank, and one front oblique aspect. Rear aspects are excluded. Reference materials relying on photographs typically present close-up shots where range and aspect are not varied systematically.

Recently, computer-based programs for visible aircraft recognition training have been produced (Jane's Information Group, 1993a, 1993b; ADA, 1993). These programs take a more systematic approach to presenting aircraft cues. In the ADA program, incidental cues were eliminated by creating computer-automated drafting models of aircraft. The ADA program also provides multi-aspect views of many aircraft at different ranges. The Jane's testing program uses photographs in which the background has been selectively deleted.

Vehicles in CVIPlus

CVI*Plus* contained 14 combat vehicles. They are listed by type and US/nonUS categories in Table 1. Abbreviations used in the report for particular vehicles are also cited.

Table 1
Combat Vehicles in CVIPlus

Type	Vehicle (abbreviation)	US or Non US
Tanks	M1A1 Abrams (M1)	US
	T72	NonUS
	M60A3 (M60)	US
	M551 Sheridan (M551)	US
Infantry Fighting Vehicles/ Armored Personnel Carriers	M2/M3 BFV (M2)	US
	M113	US
	LAV-25 (LAV)	US
	BMP-2 (BMP)	NonUS
	BTR-80 (BTR)	NonUS
Logistics/Support	M998 HMMWV (HMMWV)	US
	M35 2.5-ton truck (M35)	US
	M814 5-ton truck (M814)	US
	ZIL 131 truck (ZIL)	NonUS
Air Defense Artillery	ZSU 23/4 (ZSU)	NonUS

Note. Versions 1.1 and 1.3B of CVI*Plus*.

CVIPlus Modules

Soldiers progress through CVI*Plus* at their own pace. Prior research (Gibson, 1947; Warnick & Smith, 1989; Whitmore, Cox, & Friel, 1968) has shown large individual differences in the ability of individuals to acquire identification skills. Group instruction, often conducted at the rate of the slowest learner, is inefficient when compared to a self-paced, computerized training program. The self-pacing capability of CVI*Plus* was designed to reduce the training time for many soldiers, allowing them to progress quickly to higher target acquisition skill levels.

The six modules in CVI*Plus* were: Thermal Basics, Vehicle Basics, ID Training, Image Library, Practice/Training Exercises, and Scored Tests. These modules transitioned soldiers from photopic to thermal views of vehicles, gave them an understanding of the dynamics of thermal imagery, trained them on vehicle types and nomenclature, allowed them to compare the unique thermal signatures of vehicles, and provided practice in discriminating vehicles. Appendix F presents typical screen displays from the modules used in the experiments described in the report.

• <u>Thermal Basics</u>: Instruction on sight controls -- brightness, contrast, polarity (white-hot and black-hot), focus, field of view, and reticle brightness.

• <u>Vehicle Basics</u>: Instruction and short exercises on vehicle names and types; comparison of visible and FLIR images.

• <u>ID Training</u>: The ID Training module was designed to demonstrate how physical, and typically the visible, features of vehicles translate into hot spots and cool surfaces as seen with FLIR imagery. The soldier selected from four vehicle pairs for comparison: M1 and T72, M60 and T72, LAV and BTR, and M2 and BMP. The soldier also was given the option of viewing the changes from visible to thermal from four different aspects: front, rear, and both flanks. Once the soldier selected the aspect and the vehicle pair to compare, a video "morph" was displayed in two boxes, placed side-by-side in the center of the screen. The video morph was simply a gradual display of a shift from a visible image of the vehicles to a white-hot FLIR image to a black-hot FLIR image and finally back to a visible image. The shifts were sequential, successive approximations from one way of viewing the vehicle to another. The morph served to highlight the relationship between physical features of the vehicles and features of the thermal signature.

• <u>Image Library</u>: The Image Library was the storehouse of all FLIR and visible images. It provided the user the flexibility of looking at all the images in the training database. Vehicles could be displayed at each of the four ranges, with day and night imagery, and in white-hot and black-hot modes. The four ranges were labeled Close-up, Range 1, Range 2, and Range 3. The exact distances were not cited in order that the imagery remain unclassified. However, the size of the vehicle image at the Close-up Range was twice as large as Range 1, three times larger than Range 2, and four times that of Range 3. The Close-Up Range provided a distinct view of the heat plant of each vehicle; the farthest range, Range 3, provided a view typical of what a gunner

would see at tactical engagement ranges. Also included were visible images of the vehicles at the Close-up Range only. From the toolbar at the top of the Image Library screen, the soldier could select either a two-vehicle display screen or a three-vehicle display screen. These side-by-side displays allowed the user to compare different vehicles or different images of the same vehicle. At the bottom, were two (or three) rows of small images that presented the eight aspects of the vehicles. These images were buttons that controlled the aspect angles of the vehicles presented in the two (or three) screen display. By clicking on these buttons, the corresponding aspect could then be displayed as one of the large images. A pull-down menu displayed the vehicle choices available. The toolbar also allowed the user to select the range, type of imagery, and polarity, when thermal imagery was selected. Labels and icons attached to each image in the center display defined the image's parameters: the vehicle, its range and aspect, the polarity setting, and whether the imagery was collected during the day or night.

- <u>Practice/Training Exercises</u>: Five practice, also called training, exercises were available. Each exercise provided immediate feedback after each trial. A trial is defined as any display where the soldier must respond by identifying a single vehicle or discriminating between two or more vehicles.

o <u>Vehicle Comparison</u> presented two images and required the soldier to indicate whether they were the same vehicle or different vehicles. The soldier responded by clicking on the "same" or "different" button and was free to change his answer until he selected the "done" button. Immediate feedback indicated whether the soldier was correct and also labeled each vehicle. Pairs of images were presented randomly from a predefined file of image pairs. Upon exercise completion, the percentage of correct responses was presented along with the number of trials and an average response time.

o <u>Aspect Matching</u> presented a random display of the eight aspects of a vehicle, and required the soldier to match each aspect to an icon depicting that aspect. The matches could be changed until the soldier was satisfied with all selections. The exercise had to be repeated until all aspects for a vehicle were matched correctly. Vehicles were presented randomly from a file of image sets. The number of correct matches was presented as feedback, and a red X placed over the incorrectly matched images.

o <u>Vehicle Matching</u> required the soldier to determine which of four images was the same vehicle as the vehicle in the center of the screen. When a correct match was made, the vehicle's name was presented in green in the upper right hand corner of both the center display and the matched display. When an incorrect response was made, a message reading "incorrect" appeared below the center image, and the correct image was highlighted by a green box drawn around it. The incorrectly selected image was surrounded by a red box, and the vehicle names printed in the upper right hand corners of the incorrect, correct, and sample images. Upon completion of the exercise, the percentage of correct matches and total number of trials was presented. A soldier could change a response once made. Sets of images were presented in randomized order from a file of image sets.

o Signature Challenge was a timed exercise, requiring the soldier to name a single vehicle displayed from a set of vehicles presented within the exercise. The soldier could be required to discriminate between as few as two or as many as nine vehicles. Responses/choices were limited to the vehicles within the training set. Thus if two vehicles were trained in a set, only the names for these two vehicles were presented to the soldier. If six vehicles were trained, six names were presented as options. After each response, the soldier received color-coded feedback in the selected vehicle-name button, that is, confirmatory feedback if correct (the button pressed remained green); corrective feedback if incorrect (the selected button turned red and the correct response button turned green). Gates or criteria were established for each vehicle set, so a soldier could not progress until a pre-specified degree of skill was achieved. This criterion was defined in terms of the percentage of images identified correctly. The exposure time for vehicles within a set could be modified, as well as the criterion to progress. A soldier could not change a response once it was made. Images were presented randomly within a set. Sequential sets of vehicles could be designed to meet training needs. Vehicles were presented from a file of image sets. The time of exposure and the gates were also established in this file for each set.

o Vehicle ID required the soldier to name the displayed vehicle and to indicate its aspect. The soldier could review all responses and change them until satisfied. There was no time requirement. Text feedback regarding the correctness of these responses was given after each trial. If the correct buttons were selected, they turned green when the soldier indicated he was done examining the image. If an incorrect selection was made, the selected button turned red and the correct button turned green. When the soldier erred in identifying the vehicle or aspect, the image of the vehicle he selected was shown beside the original display, providing corrective, side-by-side visual feedback as well. Images were presented randomly from the preset file. Summary feedback, percent correct, was provided for each dimension (i.e., vehicle identification and aspect). The version of Vehicle ID used in Experiment 1 required the soldier to indicate whether the vehicle was US/nonUS, tracked/wheeled, and its type (e.g., tank, logistics, armored personnel carrier) as well. The version in Experiments 2 and 3 required ID and aspect responses only. Vehicle ID differed in some major ways from Signature Challenge. There was no time requirement in Vehicle ID; soldiers had to indicate the vehicle's aspect as well as identify it; the names of all vehicles were response options; soldiers could change their responses; and visual corrective feedback was presented on the trials where errors were made.

• Scored Tests: These tests were scored versions of each practice exercise. Vehicle imagery presented in the scored tests could be tailored to the trainer's or user's requirements; they were not necessarily a mirror image of the training exercises. Neither trial nor summary feedback was given.

Text input files controlled the images presented in the practice exercises and tests. In addition, pre- and posttests developed specifically for the CVI*Plus* research were used. The pre- and posttests were identical. Each had a FLIR and a visible (photopic) component; each consisted of 30 displays. Three images of ten vehicles were presented. All vehicles were displayed at the closest range within the program. A flank, front or frontal oblique aspect, and rear or rear oblique aspect were presented for each vehicle. The ten vehicles in these tests were:

M1, M60, T72, M2, LAV, BTR, ZSU, BMP, HMMWV, and M113. The names of all 14 vehicles in the program were displayed as choices on the pre- and posttests. More information on these tests is in Table A-1.

Throughout this report, several key terms are used to describe the CVI*Plus* training and test features. The definitions of these terms are:

Training Exercise: A submodule within the CVI*Plus* Practice module, such as Signature Challenge or Vehicle ID, that trained vehicle identification skills. On training exercises, feedback was provided to the soldier after each response (trial).

Test: A submodule within the CVI*Plus* Scored Test module, such as Signature Challenge or Vehicle ID, or a pretest or a posttest that assessed vehicle identification skills. On tests, no feedback was provided.

Vehicle Set: Applies to the tests and the training exercises. Refers to the group of vehicles and the corresponding images used in a training exercise or test. In Signature Challenge there could be several vehicle sets. Signature Challenge allowed the soldier to train on one vehicle set and then progress to other sets after achieving a specified level of proficiency on the prior set. Sequential testing with Signature Challenge was also possible, but no feedback was given and no criterion was established for progressing to the next set.

Session: Applies to Signature Challenge training exercises only. Refers to the number of times the soldier was automatically required to repeat a vehicle set in order to reach the training criterion established for that set. [In Experiment 2, there were multiple sessions with the Vehicle ID training exercise, as the research staff imposed a criterion for exiting the vehicle set and monitored soldier progression.]

Trial: Presentation of any vehicle display, in an exercise or a test, that required the soldier to identify the vehicle presented or to discriminate between multiple vehicles.

The Training Experiments

Purpose

Three training experiments were conducted. The purpose of each was as follows:

- Experiment 1 examined the feasibility and effectiveness of different part-task training strategies. The intent was to obtain insights into the best strategy for training a large number of vehicles, as it was assumed that all vehicles could not be trained in a single period of instruction.

• Experiment 2 compared the effectiveness of fixed-pace training exercises with knowledge of results feedback to self-paced training exercises with enhanced knowledge of results (i.e., corrective visual feedback). This experiment compared the Signature Challenge and Vehicle ID exercise formats.

• Experiment 3 examined the effects of training soldiers to identify vehicles at near versus far ranges. Near ranges were the Close-up range and Range 1; far ranges were Ranges 2 and 3.

The paradigm for the three experiments is outlined in Table 2. Group instruction was used to teach the thermal signatures of vehicles. The soldiers were "on the computer" for all the other events cited in Table 2 and progressed at their own rate. However, Signature Challenge exercises imposed an upper limit on their time to respond.

Table 2
Experimental Paradigm

Sequence	Events	Description
1	Pretest with FLIR images	Same in all experiments.
2	Pretest with visible images	Same in all experiments.
3	Group instruction	-Instruction varied with experiment. -Group size corresponded to the 7 computers available. Instruction repeated to several groups to accommodate entire sample size. -Instruction conducted via overhead projection of the CVI*Plus* imagery in the Image Library.
4	Practice/Training Exercises	-Exercises varied with the experiment. -Variation corresponded to experimental conditions. -Random assignment of soldiers to experimental conditions. -Vehicle sets varied. -Soldiers proceeded at own rate.
5	Transfer tests	-Transfer tests varied with the experiment. -Soldiers proceeded at own rate.
6	Posttest with FLIR images	Same in all experiments.
7	Posttest with visible images	Same in all experiments.

Group Instruction

As indicated in Table 2, common to each experiment was group instruction on thermal signatures. All group instruction was conducted by a military instructor from the Bradley Leader Course at the Infantry School. This instructor taught the combat vehicle identification phase of the Bradley Leader Course. He had extensive knowledge of the visual and thermal

characteristics of combat vehicles as well as their physical properties and battlefield capabilities. He used the Image Library screen for all group instruction.

Thermal cues presented by the instructor were consistent with Biederman's (1987) theory of recognition-by-components. The instructor focused on the hot-spot components seen from most aspects and most distances. We developed a framework in which to present these thermal cues, which we called S^4HEET. S^4HEET stands for size, shape, suspension, symmetry, hull, engine, exhaust, and turret. This format is described in more detail in Table 3. Only the factors relevant to a given vehicle and its aspect were presented during the instruction. Cues were taught from the bottom of the vehicle up -- suspension to turret.

Instruction on thermal signatures characteristic of a class of vehicles preceded that on specific vehicles. This general instruction, for instance, stressed that main battle tanks typically have the engine in the rear and have very hot suspensions, because the weight of the vehicle on the road wheels produces considerable frictional heat. The T-series of Soviet tanks (T72, T62, T55) has the exhaust on the left flank toward the rear. For armored personnel carriers (APCs), the engine is typically in the front right of the vehicle, and the exhaust is toward the front of the right flank. The instructional strategy stressed recognizing or distinguishing what O'Kane, Biederman, Cooper, and Nystrom (1997) called "basic level" classes of objects first. The instructor then pointed out features that characterized each vehicle and distinguished it from similar vehicles. Some of these were what O'Kane et al. (1997) and Biederman (1987) termed viewpoint invariant features. Others were features that could be seen from many viewpoints (i.e., vehicle aspects), but not all.

Table 3
S⁴HEET Framework for Thermal Cues

Framework	Description	Examples
S: Size[a]	Size of vehicle relative to other vehicles in its class.	M60 is a large tank with a high profile. T72 is a small tank, with a flat silhouette.
S: Shape	Overall shape of the vehicle.	M113 is rectangular, often referred to as a toaster or a brick on tracks.
S: Suspension	Heat signature of tracks and wheels; nature of skirts, sprockets, etc.	M1 has hot suspension because of friction from heavy road wheels. The cut-out for the sprocket connects heat from the suspension with heat from the exhaust. Skirts are cooler at night than the suspension when tank is running. T72 has rubber/vinyl skirts that hide heat at night, but heat up during day.
S: Symmetry of Thermal Signature	Whether both flanks present the same thermal signature; whether right and left sides of the front are mirror images of each other, and whether the right and left sides of the rear are mirror images.	Both flanks of the M1 are basically symmetric, as are the rear and front. M2 flanks are not symmetric as the thermal signature from the engine and exhaust appears on the right only. Front is asymmetric because of engine location. Rear hull is symmetric, but turret and upper part of hull is asymmetric because of exhaust signature and TOW launcher.
H: Hull	Shape and thermal characteristics of the hull.	Flat top on T72 hull quite visible from the flanks.
E: Engine	Location of engine.	M1, T72, and M60 engines are centered in the rear. M2, BMP, and LAV engines are on front right.
E: Exhaust	Location of engine and direction of exhaust.	M1 exhaust in the rear. Turbine engine presents a very hot signature; exhaust directed downwards and forms a plume. T72 exhaust is on the left side toward rear of hull; exhaust is directed downward.
T: Turret	Thermal characteristics of the turret.	M1 has needle nosed turret with angular sides. Storage boxes on sides present a cool signature. M60 turret is rounded and big. Large cupola on right is visible at close ranges.

[a] Size is best illustrated when there is a standard for comparison (such as a 6 ft. tall man) or when two or more vehicles are photographed simultaneously for comparison purposes.

Experiment 1: Part-Task Training

Experiment 1: Background

Learning to identify a large number of vehicles is a very difficult task. This skill cannot be achieved in a single block of instruction. Some form of part-task training is necessary in order to master all vehicles, the total task. This rather pragmatic question was raised in earlier research on visual aircraft recognition/identification training programs (Gibson, 1947; Whitmore et al., 1968). There are two major instructional design issues: the criteria to use in dividing the total pool of vehicles/images and how to sequence the training on these sets. The first issue has typically been addressed in terms of similarity of vehicle (or aircraft) images. Gibson raised the question of whether similar aircraft should be trained together initially for contrast or trained separately to minimize initial interference. Based on a preliminary experiment where no significant differences were found between training on similar and dissimilar aircraft, Gibson concluded that neither hypothesis was confirmed. Whitmore et al. (1968) divided aircraft by similarity. In the visual vehicle identification training experiments at ARI-Ft. Hood (Warnick & Kubala, 1979; Warnick & Smith, 1989), vehicles were not grouped by similarity. Instead, sets of five vehicles were generated, composed of at least one main battle tank, at least one small tank or tank-like vehicle, one square or box-shaped vehicle, and another small vehicle.

In Experiment 1, we addressed the second instructional design issue: how to create an effective training sequence with several vehicle sets. The focus was on the training sequence rather than the similarity or dissimilarity of vehicles within sets. No formal assessment of vehicle similarity was made. However, the vehicle sets established were based on vehicle type, when possible, that is, tanks, tracked APCs, wheeled APCs. By default, therefore, vehicles were grouped by some degree of similarity, as vehicles within these classes share some thermal signatures.

Part-Task Training Strategies

Experiment 1 compared different part-task training strategies or schedules. The critical question in part-task training is how well training on components transfers to the whole task. Proctor and Dutta (1995) distinguished between segmented and fractionated training schedules. Segmented schedules refer to part-task schedules where a task is split into components along spatial or temporal dimensions. Fractionated schedules refer to breaking a task into components that are typically performed concurrently. Fractionated schedules are relevant to the task of combat vehicle identification, as we considered the whole task to be that of identifying all images in the program, with task components referring to subsets of images from the total pool.

Proctor and Dutta (1995) outlined three types of fractionated training schedules: repetitive-part, pure-part, and progressive-part. These are illustrated in Figure 1-1.

Repetitive-Part Schedule

Step 1	*Step 2*	*Whole Task*
Part A	Part A	Part A
	Part B	Part B
		Part C

Pure-Part Schedule

Step 1	*Step 2*	*Step 3*	*Whole Task*
Part A			Part A
	Part B		Part B
		Part C	Part C

Progressive-Part Schedule

Step 1	*Step 2*	*Step 3*	*Step 4*	*Whole Task*
Part A		Part A		Part A
	Part B	Part B		Part B
			Part C	Part C

Figure 1-1. Experiment 1: Illustration of fractionated training schedules (based on Proctor & Dutta, 1995)

A variation to these part-task paradigms was required to apply them to vehicle identification. In this experiment, part-task training schedules were used during the CVI*Plus* training/practice exercises to help soldiers learn to discriminate one vehicle from other vehicles. For these exercises, a "part" was defined to be imagery of at least two vehicles. This imagery also entailed several views or images of each vehicle; for example, all eight aspects under a specific condition such as night, white-hot. An example of a pure-part strategy with six vehicles is given in Figure 1-2. Another variation of a pure-part strategy with six vehicles would be to train three at a time, again culminating with all six vehicles.

Step 1	*Step 2*	*Step 3*	*All Vehicles*
Vehicle A Vehicle B			Vehicle A Vehicle B
	Vehicle C Vehicle D		Vehicle C Vehicle D
		Vehicle E Vehicle F	Vehicle E Vehicle F

Figure 1-2. Example of a pure-part training schedule with six vehicles.

Whitmore et al. (1968) used a variation of a pure-part schedule. They viewed aircraft recognition training as a form of discrimination learning. Judges classified 16 aircraft on the basis of similarity. There were four sets of aircraft with four planes per set. After an initial session on determining the features that distinguished pairs of aircraft, the part-task training on naming aircraft began. Images of aircraft within the first set were presented one at a time, and soldiers had to name each aircraft. Progression to the next set was based on a criterion of 80% correct. These training sessions were 50 minutes in length. Each session ended with a test on all 16 aircraft, even though soldiers had not necessarily received training on all the aircraft at that point. Feedback on test scores was provided. On average, a total of 16 sessions was required to meet the overall criterion of 95% correct for all aircraft, an average of one hour of training per aircraft. This strategy conforms to the pure-part schedule depicted in Figure 1-1, except for the all aircraft test given after each part-task session. The test at the end of each training session provided a review and practice on previously learned aircraft, plus additional familiarity on aircraft in the next-to-be-trained sets, and could have influenced the final test results. The authors acknowledged that the contribution of each component of the training program toward the final performance could not be determined.

A pure-part schedule was also used in the vehicle identification research program at ARI-Ft. Hood (Smith, Heuckeroth, Warnick, & Essig, 1980; Warnick & Kubala, 1979; Warnick & Smith, 1989). In these training experiments, typically 25 vehicles were divided into sets of 5 vehicles each. Five aspects of each vehicle composed a vehicle set (no rear or rear oblique views). In reality, a vehicle set of 25 displays was divided into a series of 5 blocks with 5 displays each. In each block, each vehicle was presented at one of the available aspects. Progression to the next set of 5 vehicles was gated. Before progressing to the next set of vehicles, however, a test on vehicles in the set just trained was administered. A test on all vehicles was given at the end of all part-task training and testing. The tests used a sample of the training imagery; not all aspects were presented.

Gibson (1947) used both pure-part and whole-part training strategies in his aircraft recognition experiments. None of these earlier efforts compared the relative merits of different part-task training strategies.

In summarizing a series of experiments on skill learning and transfer, Lintern (1989) indicated that a part-task strategy based on a systematic decomposition of a complex task showed substantial and persistent enhancement in transfer. However, the findings also indicated that in decomposing tasks, critical integration skills may not be trained; components that offer no substantive learning challenge may be selected for intensive training; and learned high levels of skill do not always develop resistance to interference from additional loads.

Choice of Part-Task Training Strategy for Experiment 1

The original intent of Experiment 1 was to compare two part-task training strategies: a pure-part strategy and a progressive-part strategy (Proctor & Dutta, 1995). Whole-task training was to follow the part-task training. The whole-task was to identify eight combat vehicles. The pure-part strategy was to be training on pairs of vehicles, for a total of four pairs, concluding

with training on all vehicles. The progressive-part strategy also involved training on vehicle pairs. However, after training on the first two pairs, training on these four vehicles was to be conducted. Then the next pair of vehicles was to be trained, followed by training on the six vehicles. This progressive strategy would continue, culminating with all vehicles. At a minimum, all eight aspects of each vehicle would be presented during training.

We believed the progressive-part strategy would provide an indication of the value of integrating skills as additional vehicles are gradually added to the program, but without intensive, repeated training on the vehicles at the beginning of the training sequence. This expectation was consistent with the insertion of tests at the end of part-task training sessions to provide a review of all aircraft being trained (Whitmore et al., 1968). We felt that interference between vehicle sets on the final task was likely if there had been no integration of vehicle sets in training prior to that point (i.e., pure-part schedule). In fact, Smith et al. (1980) found a significant degradation from the average identification score of 83% correct on the part-task tests to the final test score of 39% correct on all vehicles. Most likely, confusions between vehicles in the separately-trained sets produced the lower score on the final test. We expected, therefore, that the progressive-part strategy would be better than the pure-part.

However, due to unanticipated limitations in the prototype CVI*Plus* software, the progressive-part strategy could not be examined. IconAuthor™ allowed only an increase in the number of vehicles displayed in a sequential training strategy, not a decrease. Therefore, two pure-part strategies, which varied the number of vehicles in a set, were examined instead. The initial paired strategy (four sets of two vehicles) was retained as planned. The second strategy had three vehicle sets: one set with two vehicles and two sets with three vehicles. Within each strategy, the vehicle sequence was counterbalanced to control for order effects. It was anticipated that vehicles trained in the last set would have less interference from other vehicles during training on the whole-task vehicle set.

The three-vehicle set strategy was expected to produce higher scores when all vehicles were assessed, since it provided more integration than the pair strategy. Yet, it was possible that going from two to three vehicles would not provide substantially more integration. On the other hand, increasing the number of vehicles to four or more could have resulted in a considerable increase in training load, creating a quite difficult task. One disadvantage of larger vehicle sets is that the number of images presented can increase substantially. For example, when all aspects of a vehicle are shown during training, the addition of a one more vehicle adds at least eight displays; two vehicles add at least 16 displays. As vehicles are added, training time increases (Smith, Heuckeroth, Warnick, & Essig, 1989).

The revised design did not include the repetitive-part schedule. A repetitive schedule provides more training on the vehicles presented at the first of the sequence than those at the last. Such a strategy would have been applicable if there had been prior information on the relative difficulties of vehicles for the training audience. Then the most difficult vehicles could have been presented first. Since this information was unavailable and the size of the soldier sample did not allow for counterbalancing all orders of vehicles, the repetitive-part schedule was not

included. In addition, there was concern that when vehicles were added one-by-one, soldiers might score well by simply noting that an image had not been presented before and therefore it must be the new vehicle on the list, regardless of its identity. In other words, the task would become one of discrimination, not of naming or identification.

Two transfer conditions were also investigated. These were included to provide insights into the relative difficulty of identifying vehicles with thermal imagery under different conditions. The range was increased in one condition. In the other, day thermal imagery was used instead of night thermal.

Experiment 1: Purpose

The primary questions addressed in Experiment 1 were:

- Which pure part-task schedule is the most effective means of training vehicle identification?

- Which part-task schedule produces the higher transfer to other exemplars of the vehicles? Transfer to the same vehicles at different ranges with imagery taken at the same time of day and to the same vehicles at the same range but with daytime instead of night imagery was examined.

- Is vehicle identification skill greater for variations in range or variations in the time of day, regardless of the part-task training schedule?

The secondary questions of interest were:

- Does thermal imagery training transfer to visible imagery regardless of the training condition?

- Does thermal imagery training transfer to vehicles not trained regardless of the training condition?

- What vehicles tend to be confused with each other during training and testing?

It was expected that the three-vehicle set strategy would produce higher scores when all vehicles were assessed, since it provided more integration than the pair strategy. It was also expected that identification scores for vehicles trained last in the sequence would show less interference from other vehicles when all vehicles were assessed.

Based on Osgood's (1949) transfer surface, it was expected that soldiers would have more difficulty transferring to the day white-hot, thermal imagery than transferring to the next distance in black-hot, night imagery. The similarity of the thermal signatures of vehicles is higher between ranges than between day and night images. With range, cues are likely to merge

or fade away, but new and distinct cues or shapes are not visible. On the other hand, day thermal imagery can vary substantially from that at night. Vehicle structures can appear quite changed, as the relative temperature differences vary over the diurnal cycle. For example, a cool turret appears dark at night and is difficult to see. During a sunny day the turret heats up, and its shape can be quite evident. Suspension systems are typically quite hot and therefore visible at night, but during the day, their heat signature can appear cool relative to the vehicle's hull.

Experiment 1: Method

Participants

The participants were 24 soldiers from a Ranger unit being trained on the Javelin, a new, shoulder-fired, anti-tank missile system. The command launch unit of the Javelin includes a second-generation FLIR sight which is used by the gunner to acquire targets during day and night. The majority (54%) were Privates, 25% were Specialists, 13% were Sergeants or above, and two were First Lieutenants (see demographic form in Appendix A).

Their average years of service was 2.5 years. Many (46%) had one year of military experience; 37% had two to four years experience, two (8%) had six years of service. One soldier had less than one year experience; one had twelve years experience.

Antiarmor gunners made up 67% of the sample; 17% were ammunition bearers; while the others (17%) held leader positions (platoon leader, platoon sergeant, section leader). On average, the time in these duty positions was 9.35 months ($SD = 9.72$). There was little variability among soldiers, with 62% having six months or less of training in their current position. Another 17% had between 8 and 11 months in position; 21% had 20 months or more in position.

Soldiers were asked about their previous experience with FLIR imagery. The majority (92%) had no previous experience with this technology. The remaining 8% had worked with the thermal sights on either the M2 or the M1. Information on age, combat experience and Combat Training Center (CTC) experience were not obtained, as they were not included in the automated sign-on procedure for Version 1 of CVI*Plus* used in this experiment. However, as most of the soldiers were inexperienced, it is unlikely that they had been in combat or gone to a CTC.

Design

The design for Experiment 1 is outlined in Table 1-1. It follows the paradigm presented in Table 2. The sequence of events was: pretests, group instruction, individual practice/training on identifying vehicles, transfer test, and posttests. Soldiers were trained on eight vehicles: the M1, M60, T72, M551, M2, LAV, BTR, and ZSU. Soldiers were randomly assigned to one of four pure-part training schedules, which were executed using the CVI*Plus* Signature Challenge training exercises. The pure-part strategy with four sets of vehicle pairs is labeled "Diad." The pure-part strategy with two sets of three vehicles and one set of two vehicles is labeled "Triad."

Table 1-1

Design of Experiment 1 and Sequence of Events

Sequence of Events	Description			
1. Pretests	FLIR test, then Visible test.			
2. Group Instruction on Vehicles and Vehicle Cues	Overhead projection using Image Library. White-hot, night images. -Vehicles were: M1, T72, M60, M551, M2, LAV, BTR, ZSU -Each vehicle presented at the close-up range. Side-by-side comparison with visible image on the 2-display screen. Each aspect angle displayed. -Summary: Comparisons of night images on the 3-display screen: M1, T72 and M60 (tanks) M2, M551, and ZSU (other tracked vehicles) BTR, LAV and M2 (APCs)			
3. FLIR ID Training with Signature Challenge	Random assignment of soldiers to Signature Challenge conditions. - Pass criterion for each vehicle set was no more than 4 errors. - Exposure time longer when more images presented. - 80% of the imagery was white-hot, night at Range 1. See text for details.			
a. Part-Task	**Diad A** ($n = 6$) ① M1 & T72 ② M60 & M551 ③ M2 & ZSU ④ BTR & LAV	**Diad B** ($n = 6$) ① BTR & LAV ② M2 & ZSU ③ M60 & M551 ④ M1& T72	**Triad A** ($n = 6$) ①BTR & LAV ② M1, T72 & M60 ③ M2, ZSU & M551	**Triad B** ($n = 6$) ① BTR & LAV ② M2, ZSU & M551 ③ M1, T72, & M60
b. Whole-Task	⑤ All 8 vehicles	⑤ All 8 vehicles	④ All 8 vehicles	④ All 8 vehicles
4. Transfer Test (See Table 1-2)	Scored Vehicle ID test module. All 8 vehicles in Signature Challenge Training; total of 64 images. Test was not timed.			
5. Posttests	Same as Pretests			

Note. Images were presented randomly within each vehicle set.

Group instruction on vehicle cues. After the pretest and prior to Signature Challenge training, soldiers received group instruction on the identifying characteristics of each vehicle in the training set. This was done using the Image Library module of CVI*Plus*. The Image Library screen was projected via an overhead projection system. All instruction was conducted using the Close-up Range in the Image Library. Instruction was presented by the military instructor using the S^{4}HEET framework described previously.

The two-vehicle display in the Image Library was used to present each vehicle. The day visible image was displayed on the left side of the screen, the white-hot, night thermal on the right. Thermal cues for each vehicle were identified as each aspect angle was presented and

discussed. After all vehicles were presented, the three-display screen was used to compare the white-hot, night thermal images of the vehicles. The M1, T72 and M60 tanks were compared; the M2, M551, and ZSU; and finally the LAV, BTR, and M2. Vehicle similarities and differences were discussed, by aspect angle.

Only six computers were available for training. Therefore, the group instruction was repeated four times in order to accommodate all 24 soldiers.

Signature Challenge training exercises. The Diad training strategy involved training on two vehicles at a time, concluding with training on all vehicles: a 2, 2, 2, 2, 8 sequence. The Triad training strategy began with training on two vehicles, followed by two training sessions of three vehicles each, and concluded with training on all vehicles: a 2, 3, 3, 8 sequence. Vehicles were grouped by type within each schedule.

In the two Diad conditions, the vehicles were grouped by type (as shown in Table 1-1): two main battle tanks currently in use - M2 and T72; the two other tanks - M60 and M551; the other two tracked vehicles - M2 and ZSU; the two wheeled APCs - BTR and LAV. Within the two Triad conditions, the vehicles were also grouped by type and whether they were tracked or wheeled: three main battle tanks - M2, T72 and M60; the other three tracked vehicles - M551, M2, and ZSU; the two wheeled APCs - BTR and LAV. The part-task sequence was reversed within the Diad and Triad conditions to counterbalance for order effects.

Ten images of each vehicle were presented during the part-task sessions. All eight aspects of each vehicle were displayed at Range 1 in night, white-hot imagery. Two additional images were shown, both white-hot: the front at Range 2 in night imagery and the front in day thermal imagery at Range 1. These presentations gave soldiers some exposure to the transfer conditions, but not with the aspect on which transfer was tested. Thus there were 20 displays when two vehicles were paired; 30 displays when three vehicles were trained. In the final, whole-task session, where all vehicles were presented, only 8 images of each vehicle (i.e., each aspect at Range 1, night, white-hot) were used, for a total of 64 images.

The criterion for going to the next set of vehicles, as well as for passing the final set, was no more than 4 errors. Pass rates required were: at least 79% correct for two vehicles, 85% for 3 vehicles, and 93% for all eight vehicles. Soldiers had five seconds to respond when two vehicles were presented, eight seconds for three vehicles, and ten seconds for all eight vehicles. If they did not respond within the time limit, an incorrect response was recorded. Soldiers clicked on the name of the vehicle being displayed; only the names of the vehicles being compared were displayed.

Testing for transfer. The transfer test was administered with the Vehicle ID test. These conditions examined the ability of the soldiers to transfer identification skills to the next range and to day thermal imagery. Three different aspects were used for each of these two conditions, as shown in Table 1-2. For the range condition, the left flank, left rear oblique, and right front oblique were presented at Range 2. For the day thermal condition, the right flank, right rear oblique, and left front oblique were presented, but at the same range used in Signature Challenge

training. The front and rear views were the same as those in Signature Challenge training (night, Range 1, white hot); no transfer was involved. Thus the Vehicle ID test had 64 displays, eight aspects of each of the eight vehicles.

Table 1-2
Experiment 1: Format of Vehicle ID Test

No Transfer: Night Range 1 White Hot		**Transfer:** *Range 2* Night White Hot			**Transfer:** *Day Thermal* Range 1 White Hot		
Aspect Angle							
Front	Rear	Left Flank	Left Rear Oblique	Right Front Oblique	Right Flank	Right Rear Oblique	Left Front Oblique

Soldiers had to make five responses on the test. As with all other training exercises and tests, they had to identify the vehicle. But on the Vehicle ID test, they also had to determine the vehicle's aspect, whether it was US or nonUS, whether it was tracked or wheeled, and its type (e.g., tank, APC, logistics, ADA). Only the identification and aspect responses were analyzed. The names of all 14 vehicles in the program were displayed and were response options for the soldier. Soldiers were not told that they would be tested on only the eight vehicles used in Signature Challenge training. This test was not timed.

Pretests and posttests. The pretests and posttests were identical. Each had a FLIR and a visible component, consisting of three images of ten vehicles. These 30 images were presented randomly. All vehicles were displayed at the closest range. A flank, a front view including an oblique, and a rear view including an oblique were presented for each vehicle. The ten vehicles in these tests were: M1, M60, T72, M2, LAV, BTR, ZSU, BMP, HMMWV, and M113. Of these ten vehicles, seven had been trained in Experiment 1 and three (BMP, HMMWV, and M113) had not. The names of all 14 vehicles in the program were displayed as choices. Soldiers were not timed. Table A-1 provides additional information on the format of these tests.

Measures of Performance

The pretests, posttests, Signature Challenge training exercises, and Vehicle ID test each required soldiers to identify the vehicle displayed. Vehicle images were presented randomly within all exercises and tests. For the pre- and posttests, soldiers identified the vehicle by selecting a name from a list of all the 14 vehicles within CVI*Plus*. For Signature Challenge training, only the vehicles being displayed were possible responses. Depending on the vehicle set, this involved a choice of two, three, or eight vehicles. For the Vehicle ID test, the buttons for all 14 vehicles in CVI*Plus* were automatically displayed as response options. Scores on both vehicle identification and aspect were computed for Vehicle ID.

Response times to each trial were recorded on each exercise and test. The number of sessions to criterion on Signature Challenge training was recorded.

Procedures and Instruments

Version 1.0 of CVI*Plus* was used. The pre- and posttests, the group instruction with the Image Library instruction, the Signature Challenge training, and the Vehicle ID transfer test were all administered using a Toshiba Tecra CDT 520 laptop computer. The computer had a 1024 x 768 active matrix, color, 12 in. diagonal screen. Except for the Image Library instruction, all training and testing progressed at the soldier's own rate.

Experiment 1: Results

Program Completion

Due to time constraints, all soldiers were unable to complete all phases of training. To insure that all soldiers took the posttests, it was necessary to have some terminate parts of the program. In particular, everyone did not train to criterion on the final vehicle set in Signature Challenge that had all eight vehicles. Also, the criterion for passing the final vehicle set in Signature Challenge was changed from 93% to 90%. This change was made to expedite training. We considered 90% a high criterion; soldiers had to get at least 58 of the 64 vehicles correct. Only six soldiers took the Vehicle ID transfer test; four completed it. One soldier did not take the posttests. Table 1-3 shows the status of soldiers at each phase of training.

Table 1-3
Experiment 1. Number of Soldiers Completing Each Training Exercise and Test

Status on Training Exercises and Tests	Diad A	Diad B	Triad A	Triad B	Total
Completed Pretests	6	6	6	6	24
Completed Each Vehicle Set Prior to the Last Set in Signature Challenge	6	6	6	6	24
Completed First Session of Last Vehicle Set in Signature Challenge	6	6	6	5	23
Met Criterion on Last Vehicle Set in Signature Challenge[a]	2	3	2	3	10
Took Vehicle ID Test	2	1	1	2	6
Completed Posttests	6	5	6	5	22
Valid Posttests[b]	4	5	5	3	17

[a] Those who were unable to reach criterion on this set did not stop at one session. The number of sessions ranged from 1 to 11.

[b] Imagery from the Vehicle ID test corrupted the posttest imagery. Only soldiers who did not take the Vehicle ID test had valid posttests.

The original plan was to use number of sessions to criterion on the final Signature Challenge vehicle set to assess the effectiveness of the different pure-part training strategies. However, this was not possible, since practice on that set had to be terminated. Scores for vehicles on the first session of this final set were used as the criterion measure instead. The number of training trials required to reach criterion on each of the part-task vehicle sets was also used as an index of training effectiveness.

Training and Testing Time

Times for each major training and testing event are shown in Figure 1-3. Times for the pre- and posttests, group instruction, Signature Challenge exercises, and the transfer test are shown. Administrative and computer down-times were excluded in order to show only the time spent in actual training. The mean times for the Diad and Triad conditions are displayed, as well as the times for the fastest and slowest individuals in each condition. Time for the transfer test is not shown as few individuals were able to complete it. Group instruction time was 55 minutes.

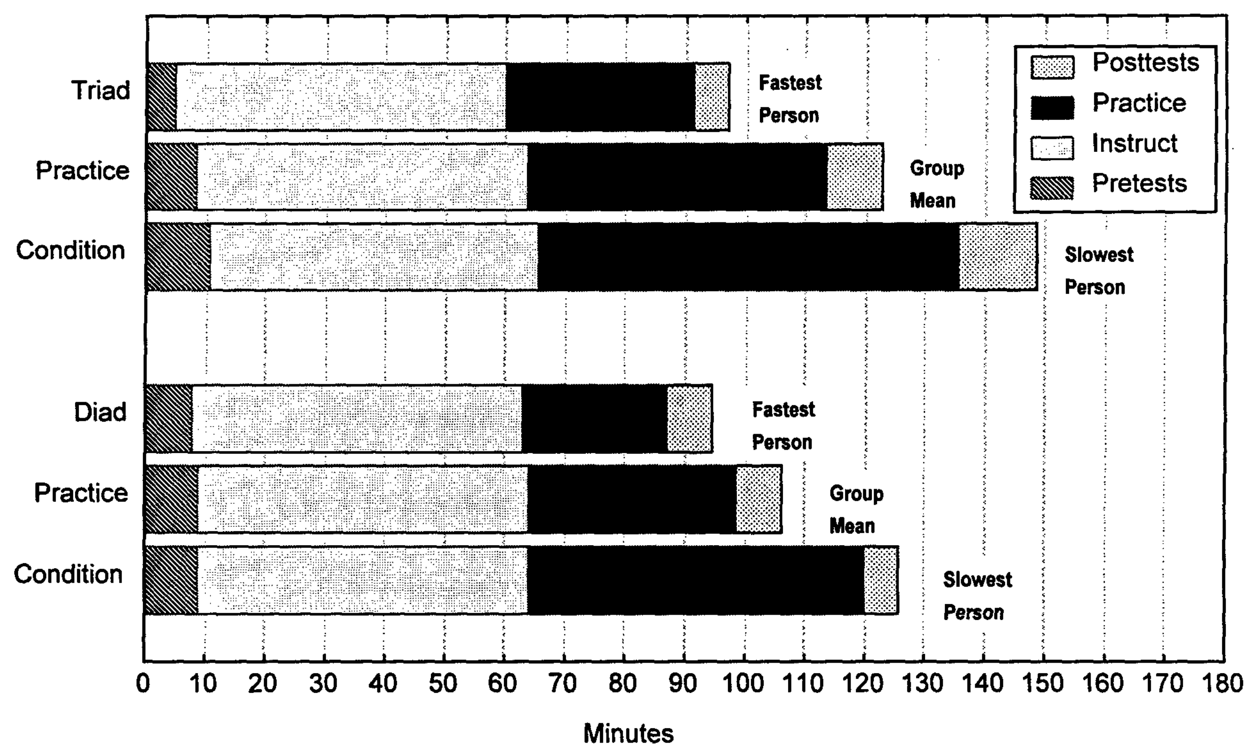

Figure 1-3. Experiment 1: Training and testing times for Diad and Triad conditions (group means and fastest and slowest individuals).

Overall, soldiers in the Triad condition required 15 minutes more than soldiers in the Diad condition. But the fastest individuals in both conditions had comparable times. On the other hand, the slowest individual in the Triad condition required 25 minutes more than the

slowest individual in the Diad condition. Table B-1 provides a further breakdown of the times. If time constraints had not occurred, the cumulative times would have been greater, as all soldiers would have taken the transfer test and been able to reach criterion on the training exercises.

Signature Challenge Training Exercises

All soldiers met criterion on every part-task vehicle set within Signature Challenge. Because time constraints made it impossible for all soldiers to reach criterion on the final set with all eight vehicles, differences between the part-task conditions were assessed by vehicle identification scores on the first session. Number of sessions to criterion on the whole task, however, could not be used to assess training differences.

The mean number of trials to criterion across all sessions within each vehicle set during the part-task training is in Table 1-4. The term "trial" refers to a single vehicle display in a vehicle set. As no significant differences occurred between the A and B variants of the Diad and Triad conditions, both variants were then combined for statistical comparisons. The average number of trials per vehicle was used to assess the relative effectiveness of the two training strategies. Soldiers in the Triad condition required 1.6 more trials than those in the Diad condition, $F(1, 22) = 9.74, p < .0049$. Soldiers in both conditions required more trials than the minimum of 10 per vehicle. Of interest as well is that the variability in performance among soldiers was greater, but not significantly so, in the Triad condition.

Table 1-4
Experiment 1: Mean Number of Trials to Meet Criterion on Signature Challenge Part-Task Vehicle Sets

Vehicle Set	Training Condition	
	Diad *M* (*SD*)	Triad *M* (*SD*)
M1 - T72	41.60 (26.23)	-----
M60 - M551	70.00 (56.89)	-----
M2 - ZSU	31.66 (15.86)	-----
BTR - LAV	66.67 (33.39)	58.33 (43.86)
M1 - T72 - M60	-----	127.50 (74.73)
M2 - M551 - ZSU	-----	162.50 (76.17)
Total # of Trials	210.00 (72.61)	348.33 (135.30)
# Trials per Vehicle	26.25 (9.08)	43.54 (16.91)

Note. Means based on 12 soldiers in Diad condition; 12 soldiers in Triad condition

Some soldiers required repeated sessions before reaching criterion on a vehicle set. For instance, 36% of the time soldiers required at least three consecutive sessions to reach criterion. Fifteen percent of the time soldiers required six or more sessions, up to 10, to pass. (Table B-2 shows the number of sessions required to pass each vehicle set.)

Mean scores on the first session of the final set with all eight vehicles are in Table 1-5. Scores are displayed for each training condition and each vehicle.

Table 1-5
Experiment 1: Mean Identification Scores (% correct) on the First Session of the Signature Challenge Whole-Task Vehicle Set as a Function of Vehicle and Training Condition

Training Condition	Vehicle								
	M1	T72	M60	M551	M2	ZSU	LAV	BTR	All Vehicles
Diad	79	67	69	39	77	42	71	78	65
Triad	77	62	78	54	73	46	52	57	63

Note. Soldier scores were the percent correct across the eight aspects. Mean scores were based on 12 soldiers in the Diad condition; 11 in Triad.

A 2 x 8 (Diad-Triad x vehicle) ANOVA was conducted with repeated measures on the last factor. There was a significant difference among vehicles, $F(7, 147) = 10.17, p < .0000$, but there was also an interaction between the Signature Challenge training conditions and vehicle, $F(7, 147) = 2.57, p < .0157$. The difference between Diad and Triad scores was less than 10 points for all but three vehicles. Diad scores were higher for the BTR and LAV, a difference of 21% and 19%, respectively; while Triad scores were higher for the M551, a difference of 15%. The main effect for vehicles shows that even though soldiers achieved a high criterion during training on each subset, when all vehicles were presented, there was "cross-set" interference. The high scores achieved during the part-task training were not always maintained when all vehicles were in the set. The relatively low scores on the ZSU and M551 best demonstrate the confusion created by other vehicles.

Also of interest were possible sequence effects associated with the order of training on the vehicle sets. The clearest indication of a recency effect was that with the LAV-BTR set. In the one instance where it immediately preceded the whole-task set (Diad A), the mean percent correct was 82%. In the other three instances, where it was the first vehicle set to be trained, the mean percent correct was 57%. The significant Diad-Triad by vehicle interaction reflects, in part, the higher scores obtained on the BTR and LAV when they were the last vehicle set trained.

Which vehicles were confused with each other when soldiers were exposed to all eight vehicles on the final set? The vehicle confusion matrix for the first session on this set is in Table 1-6. The diagonal cells represent percent correct scores; the off-diagonal cells, the vehicle confusions. For example, when the M60 was presented, 10% of the time it was confused with the M551.

Also cited in Table 1-6 are the aspect angles that created vehicle confusions. When a particular aspect resulted in a high confusion with another vehicle, that aspect is cited. High confusion was defined as an error rate of at least 15% for a specific aspect. For example, when the right flank of the M1 was shown, at least 15% of the time soldiers confused it with the T72;

when the right rear oblique of the T72 was shown, at least 15% of the time the T72 was then misidentified as the M1; when the right front of the T72 was shown, it was also likely to be misidentified as the M1.

Table 1-6

Experiment 1: Vehicle Confusion Matrix from the First Session of the Signature Challenge Whole-Task Vehicle Set

Vehicle	Vehicle Response							
Displayed	M1	T72	M60	M551	M2	ZSU	LAV	BTR
M1	**78**	5 (right)	6	3	5	2	0	1
T72	10 (rt rear) (rt frnt)	**65**	7	7 (rt frnt)	3	3	3	2
M60	2	4	**73**	10 (rt frnt)	5	2	0	4 (front)
M551	8	9 (right)	13 (front) (rt frnt)	**47**	8 (rear) (rt frnt)	9 (lf rear)	2	2
M2	3	3	4	3	**75**	9 (rt rear)	1	2
ZSU	7 (right)	11 (right) (rt frnt)	4 (lf rear)	12 (rear) (rt frnt)	12 (left) (right)	**44**	2	6 (rear)
LAV	1	3	3 (front)	2	6 (rear)	9	**62**	12 (lf frnt) (left)
BTR	1	3	2	2	1	10 (rt rear)	12 (right) (rt frnt)	**68**

Note. Percent correct responses are on the diagonal; confusions on the off-diagonal. Rows may not sum to 100% as 1% of the time no response was made. N for each displayed vehicle (row) was 184 trials. Aspect angle is indicated when confusion rate for that aspect was at least 15%. Aspect refers to aspect of the vehicle displayed, e.g., when the rear of the M1 was shown, at least 15% of the time soldiers confused it with a T72.

The confusions created by the M551 and ZSU imagery are evident. Soldiers had difficulties with practically all aspects of the ZSU and the M551. Of interest as well are confusions among vehicles that had been trained in the same vehicle set. Even though the T72 and M1 were always in the same vehicle set, confusions arose between these two vehicles when soldiers had to identify all eight vehicles. This was also the case for the BTR and the LAV. The M2 was always in the same vehicle set with the ZSU. The difficulties with the ZSU have been noted. Yet the M2 was misidentified as the ZSU when the M2's right rear was displayed.

Also of interest was the relationship between scores and time to respond. These results are presented in Table B-3 for the first session of each part-task vehicle set and the all-vehicle set. Soldiers tended to respond more slowly to vehicles that were hard to identify. This relationship was particularly evident on the final vehicle set, where the vehicles with the highest scores (at least 73% - M1, M2, and M60) also had the fastest response times (approximately 3 sec). Vehicles with the lowest scores (below 48% - M551 and ZSU) had the slowest response times (3.7 sec). The correlation between the mean vehicle identification scores and mean times to respond for the final set of vehicles across both training conditions was -.99, $p < .01$.

Even though all soldiers were unable to reach the criterion of 90% correct identifications on the final vehicle set, ten did; five in the Diad condition, and five in the Triad condition. The number of sessions for these ten ranged from three to ten. Seven of the ten reached criterion in five sessions. Although most of these soldiers showed a steady increase in score from one session to the next, this was not always the case. This variation is shown in the individual learning curves, for those that met criterion and those that did not (see Figures B-1 and B-2).

Transfer of Skill

Because of time constraints, only six soldiers took the Vehicle ID test and only four, one from each condition, completed it. The transfer analyses were based on the four with complete data, and findings represent trends only. The percentage of vehicles identified correctly as a function of transfer condition is presented in Table B-4. Scores were highest (75%) where no transfer was required, but not significantly so. Scores were similar for both transfer conditions; 65% for transfer to the greater range and 63% for transfer to day thermal imagery.

Identification scores as a function of vehicle displayed are in Table B-5. Highest scores were for the BTR, M1, and M60 (75% and greater); lowest for the ZSU (41%). Also presented in Table B-5 are the aspect angles that resulted in low vehicle identification scores, scores of 50% correct or less. As would be expected, with high scores few aspects created difficulties; with low scores more aspects created difficulties.

With the Vehicle ID test, soldiers had to determine the aspect itself as well as the vehicle. Therefore aspect scores were computed. As shown in Table B-6, the two flanks were easy to determine, with aspect angle scores above 80%. The right front oblique aspect was the hardest (aspect score of 19%), being confused 40% of the time with the right flank. An examination of the right front oblique imagery at Range 2 indicated that, for some vehicles, this angle did not appear to differ greatly from that of the right flank. Therefore, the low score for this aspect may simply be an artifact resulting from a slight misorientation of the vehicle's position from a true oblique during imagery collection. Two other aspects, the front and rear, tended to be confused with each other 17% of the time.

Pretests and Posttests

Alll soldiers took the posttest. However, due to an unexpected software problem, imagery from the Vehicle ID test was accidentally included in the posttests for the soldiers who took the

Vehicle ID test. Consequently, those soldiers did not have the same posttest as the others. Complete and valid posttest data were available on 17 of the 24 soldiers.

Pre- and posttest results on vehicle identification are in Table 1-7 for each vehicle. Findings are also displayed according to whether the vehicle was included in the FLIR training. A 2 x 2 x 2 x 2 (Diad-Triad x pre-post x visible-FLIR x train-no train) ANOVA with repeated measures on the last three factors was conducted. Main effects occurred for the pre-post and visible-FLIR factors. Significant two-way interactions occurred for pre-post with visible-FLIR, pre-post with train-no train, and visible-FLIR with train-no train. There was also a three-way interaction with these three factors. Figure 1-4 illustrates the three-way interaction.

Table 1-7

Experiment 1: Mean Percentage of Vehicles Identified Correctly on the Pretests and Posttests as a Function of FLIR Training and Type of Imagery

Vehicle	Visible Imagery			FLIR Imagery		
	Pretest	Posttest	Difference: Post - Pre	Pretest	Posttest	Difference: Post - Pre
Vehicles with FLIR Training						
BTR	27	74	47	23	94	71
M60	42	80	38	20	90	70
M2	49	75	26	45	84	39
M1	78	92	14	51	82	31
T72	44	68	24	41	80	39
LAV	31	59	28	29	69	40
ZSU	18	44	26	16	65	49
Vehicles without FLIR Training						
HMMWV	84	90	6	68	76	8
M113	60	66	6	49	61	12
BMP	19	25	6	16	20	4
Trained[a]	41	70	29	32	81	49
Not Trained	54	60	6	44	52	8
Overall[bc]	45	67	22	36	72	36

Note. Vehicle means based on 3 trials (displays) per soldier; 17 soldiers.

Significant main and two-way effects:

Pre-Post: $F(1, 15) = 44.04, p < .0000$ Pre-Post x Trn-No Trn: $F(1, 15) = 39.06, p < .0000$

Vis-FLIR: $F(1, 15) = 7.51, p < .0151$ Vis-FLIR x Trn-No Trn: $F(1, 15) = 21.18, p < .0003$

[a] Pre-Post x Vis-FLIR x Trn-NoTrn: $F(1, 15) = 10.24, p < .0059$

[b] Pre-Post x Vis-FLIR: $F(1, 15) = 12.28, p < .0032$

[c] Vehicle means weighed equally in overall mean.

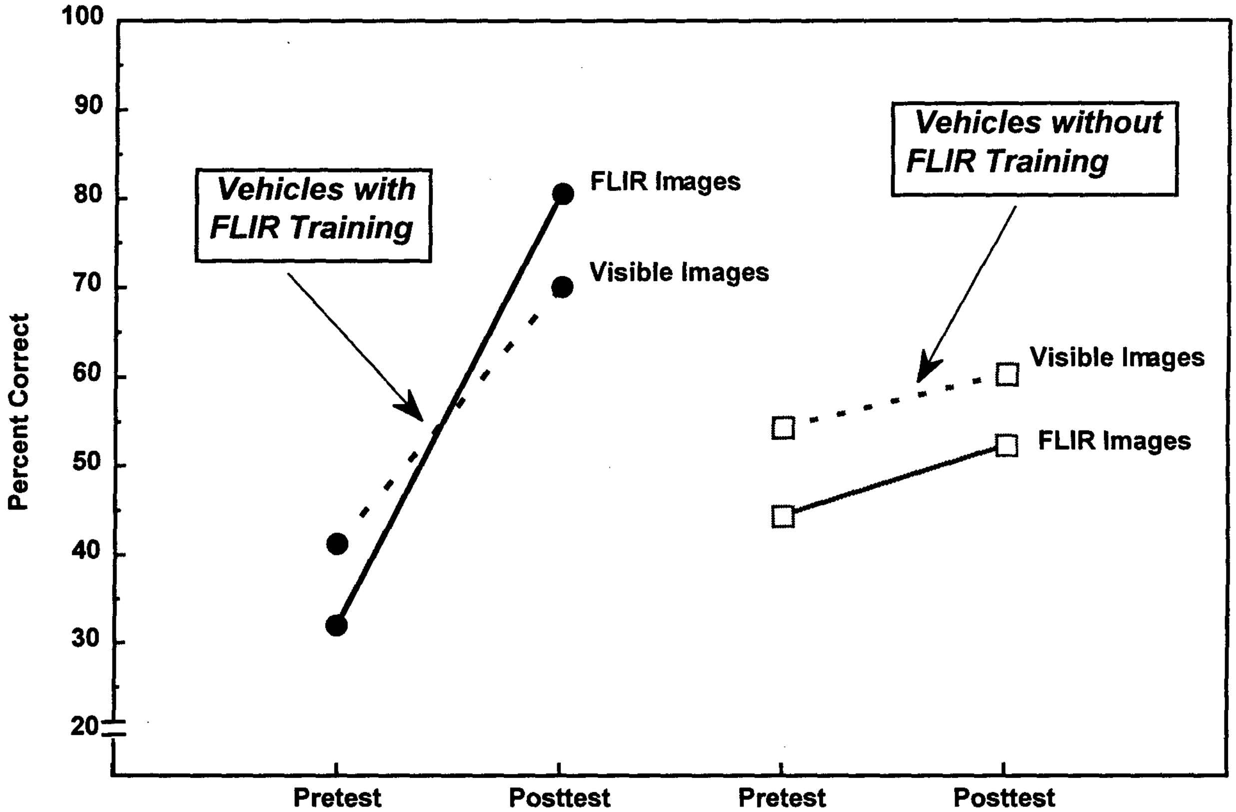

Figure 1-4. Experiment 1: Change in identification scores from pretest to posttest for FLIR and visible images as a function of FLIR training.

Posttest FLIR scores for the vehicles receiving FLIR training increased 2.5 times, from 32% to 81%. The lowest pretest scores were for the vehicles in the FLIR training. After training these vehicles had the highest posttest scores. FLIR training also transferred to the visible images of the same vehicles. Visible scores increased 1.7 times, from 41% to 70%. On the other hand, for vehicles not in the FLIR training, both visible and FLIR scores increased only slightly (6 and 8%, respectively).

Vehicle confusions on the pretests and posttests are in Appendix B. A summary of the vehicle confusions on the major tests and training exercises is presented in Table B-7. These findings illustrate again the impact of the FLIR training. Soldiers confused many vehicles on the pretests. The amount of confusion among FLIR images reduced as training progressed, and was the least at the completion of training. On the posttest, there were more confusions among the visible images than among the FLIR images.

Interestingly, of the vehicles trained, some confusions (i.e., 10% error rate) remained on the thermal image posttest. The T72 was misidentified as the M1 14% of the time; the ZSU was misidentified as the M1 16% of the time; the LAV was misidentified as the M2 14% of the time. These same confusions occurred on the visible posttest. Additional confusions existed with the

visible images: the BTR and LAV were confused with each other; the T72 was misidentified as the M551; the M60, misidentified as the M551; and the ZSU was misidentified as the M60 and M2 (see Table B-7).

Response times on the pre- and posttests were also examined. A 2 x 2 x 2 x 2 (Diad-Triad x pre-post x visible-FLIR x train-no train) ANOVA with repeated measures on the last three factors was conducted. There was a main effect for visible-FLIR, $F(1, 15) = 85.85$, $p < .0000$, but there was also a significant interaction between visible-FLIR and pre-post, $F(1, 15) = 26.80$, $p < .0001$. FLIR times dropped slightly from 10.38 sec ($SD = 3.79$) to 9.33 sec ($SD = 3.64$), while visible scores increased from 4.91 sec ($SD = 1.23$) to 6.95 sec ($SD = 2.74$).

Correlations Among Scores

Correlations among the test and training scores and number of sessions to criterion on the Signature Challenge exercises were computed. Pairwise deletion of cases was used due to the incomplete data on some of the tests.

Previous research on aircraft recognition (e.g., Whitmore et al., 1968) showed individual differences in the speed with which soldiers learned to name aircraft. Individual differences also occurred here. For example, the pretest scores correlated with many of the later training and test scores, as well as with the rate of progression. Table 1-8 presents the correlations between the pre- and posttest scores. Soldiers' initial skill in identifying vehicles with visible images correlated highly with their corresponding FLIR skills ($r = .91$), and it related highly to post visible scores. Pre-FLIR scores correlated with Post-FLIR scores despite the overall positive effects from FLIR training.

Table 1-8

Experiment 1: Correlations Between Pre- and Posttest Scores

	Pre Visible	Pre FLIR	Post Visible	Post FLIR
Pre Visible	...	.91**	.70**	.49*
Pre FLIR		...	.73**	.49*
Post Visible			...	.76**
Post FLIR				....

$^{*}p < .05$ $\quad$ $^{**}p < .01$

Pre- and posttest scores also tended to correlate with scores on the initial session of the Signature Challenge sets which had the three vehicles (see Table B-8). A similar pattern occurred with the number of sessions required to reach criterion. As shown in Table 1-9, the initial score on each Signature Challenge set also correlated highly with the number of sessions required to reach criterion on that exercise. The correlations ranged from -.57 to -.74. Soldiers who knew more initially required fewer attempts to reach criterion. Correlations with transfer scores were not computed due to the limited number of soldiers who took this test.

Table 1-9

Experiment 1: Correlations Between First Session Score and Number of Sessions to Criterion for Each Signature Challenge Vehicle Set

Vehicle Set					
M1-T72	M60-M551	M2-ZSU	BTR-LAV	M1-T72-M60	M2-M551-ZSU
-.73*	-.74*	-.74*	-.68*	-.57	-.64*

* $p < .01$

Experiment 1: Discussion

Answers to the Questions

- Which pure, part-task schedule is the most effective means of training vehicle identification?

During part-task training, soldiers in the Diad condition required significantly fewer trials to criterion on the part-task vehicle sets than those in the Triad condition. However, when all vehicles were presented, identification scores on the first training session for the Diad and Triad conditions did not differ. For both conditions, cross-set interference occurred with the all-vehicle set. Low identification scores (less than 50%) resulted for some vehicles, despite soldiers reaching the criterion of no more than four errors in each prior part-task vehicle set. There were individual differences in the time to reach criterion, suggesting that practice without remediation was not effective for those with minimal initial skill. The number of sessions to criterion on all vehicles was not addressed as time constraints prohibited all soldiers from completing this phase.

- Which part-task schedule produces the higher transfer to other exemplars of the vehicles? Transfer to the same vehicles at a different range and to daytime thermal imagery was examined.

- Is vehicle identification skill greater for variations in range or variations in the time of day, regardless of the part-task training schedule?

As only four soldiers had time for the transfer test, the findings are best described as trends. Although scores for the no transfer condition were 10% points higher than the two transfer conditions, they were not significantly higher. And identification scores for the two transfer conditions were similar. However, there simply was insufficient data to conclude that transfer conditions were similar in difficulty and reject the original expectation that transfer to day imagery would be harder than transfer to the next range.

The secondary questions of interest were:

- Does thermal imagery training transfer to visible imagery regardless of the training condition?

• Does thermal imagery training transfer to vehicles not trained regardless of the training condition?

Scores for vehicles included in the FLIR training increased from pretest to posttest. Scores remained the same for the vehicles not trained. This was the case for both visible and FLIR images. Time to respond to the FLIR images decreased on the posttest; time to respond to the visible images increased.

• What vehicles tended to be confused with each other during training and testing?

Of the vehicles trained, three confusions remained on the posttest. The T72 and the ZSU continued to be incorrectly identified as an M1 tank, and the LAV was mistaken as an M2. There were confusions between other vehicles on the pretest and during training, even on the first session of the all-vehicle set. For example, the M551 and M60 were confused with each other during training, as were the LAV and the BTR. And the ZSU tended to be confused with the other tracked vehicles. Of the vehicles not trained, the soldiers clearly did not know the BMP on the pretests and posttests. It was confused primarily with the T72, M2, ZSU, and M551.

The results that support these answers as well as findings on other aspects of Experiment 1 are discussed in more detail in the remainder of this section. The discussion section progresses from a description of the soldiers, through the learning process as evidenced by performance on the training exercises, to the culmination of instruction and training as evidenced by test results.

The Starting Point

Most of the soldiers had been in the Army for a short period, less than three years. This inexperience was also reflected in their low, pretest vehicle identification scores, both visible and thermal. Comments made to the instructor and researchers also reflected limited knowledge of combat vehicles, as some indicated no knowledge of such vehicles as the M113 or BMP. Only two of the 24 soldiers had previous experience with thermal sights. All soldiers were highly motivated and dedicated themselves to the challenge of learning the thermal signatures.

The Learning Process

Several important lessons on the training and learning processes were gleaned from this first experiment. First, soldier skill in identifying vehicles as seen through thermal sights can be increased with training. Second, learning plateaus occurred, even with trial-by-trial feedback. Third, for soldiers with weak initial vehicle skills, the number of vehicles trained at one time should be limited. Fourth, cross-set vehicle confusions occur. Training individuals to discriminate vehicles within a set to a high criterion does not mean that these vehicles are then resistant to interference from other vehicles.

How did the scores change during training? Seven of the eight vehicles trained were on the pretest. The mean pretest score for these vehicles was 32% correct. On the initial session of

the part-task vehicle sets, the mean score for these seven vehicles was 69%. This increase was probably a joint result of the group instruction and the response format. As the response format on Signature Challenge required the soldier to select only from the vehicles being presented, the probability of making chance errors declined, in contrast to the pretest where any vehicle could be selected. But the increase of 37% points cannot be attributed solely to the response format. Instruction on vehicle cues and how each vehicle looked thermally also played a role.

On the first session of the whole-task set, there was potential for cross-set confusion. Without this interference, the mean first session score probably would have been around 85%, given the exit criterion on the part-task sets. However, the mean score on the first session was 66%, indicating cross-set confusions. Some vehicles (the M551 and ZSU) were affected more than the others. There were individual differences as well. Of interest is the extent to which some vehicle scores decreased substantially for some soldiers, for example, from 100% to 25% or 50% correct for the M551; from 90% to 38% or 50% for the ZSU. These findings reinforce Lintern's (1989) statement that high levels of skill do not always mean that individuals have developed resistance to interference from additional loads.

The 66% correct score on the initial session of the whole task set indicated more learning, however, than was found by Smith et al. (1980) in a pure-part training program with photopic images. In the Smith et al. research, part-task test scores averaged 83% while whole-task scores were only 39%. In our experiment, the higher level of skill shown on the whole-task may have been due to the fewer number of vehicles trained, the self-pacing feature of CVI*Plus*, a more stringent part-task requirement for progression, or a combination of these factors.

Cross-set confusions might have been reduced more if a progressive-part task schedule had been used. Vehicle similarity within and across the sets could have also influenced the confusions on the whole-task. In a paired associate learning task, Nesbit and Yamamoto (1991) found that training on high-similarity sublists took longer than training on low-similarity sublists. Yet fewer posttest errors occurred and a smaller proportion of the posttest confusions were between high-similarity pairs. In Experiment 1, there was no deliberate attempt to place high-similarity vehicles in the same set. However, as vehicles, were in general, grouped by type, some degree of similarity did exist. Another factor affecting vehicle confusions is the vehicle aspect. Vehicles may have similar thermal signatures from some aspects, but quite different signatures from others, making it hard to place them in distinct vehicle sets. The confusion matrix from the first session of the whole-task vehicle set in Signature Challenge reflects some of these problems. For example, the right, not the left, flank of the M1 was confused with the T72. The exhaust is on the left flank of the T72, making that flank clearly distinguishable from the M1, whose exhaust is in the rear. The thermal signature of the ZSU created problems for soldiers. It is tracked and the engine is in the rear, making it easily mistaken as a tank. These findings indicate that vehicular similarity, for whatever reason (same class of vehicles, same engine or exhaust location) will create confusions. Training similar vehicles in isolation from each other to minimize initial confusion will not necessarily minimize confusions when the vehicles must eventually be distinguished from each other.

Learning plateaus occurred for the soldiers. These plateaus may have resulted from their limited initial vehicle identification skills. But the plateaus were probably also a function of the lack of summary feedback at the end of each session. Frequently, soldiers could not tell the instructor which vehicles or which aspects of the vehicles were creating problems for them. In addition, there was no provision in the prototype program for a soldier to exit a training exercise, study the vehicles in the Image Library, and then resume training. The addition of these instructional features would have lessened the likelihood of learning plateaus as well as providing soldiers with a better understanding of thermal signatures. It was clear that trial-by-trial feedback was essential, but not sufficient, for effective learning.

Findings on the whole-task vehicle set indicated that eight vehicles (64 images) placed too many cognitive demands on soldiers during a single session. The training session was too long. Soldiers became discouraged. They could easily make an accidental mistake with the computer mouse and have to repeat the session. The concentration required to respond within a fairly quick time limit also contributed to the repeated sessions on this whole-task set. Some soldiers disliked the time limit, as it put too much pressure on them during the learning process.

The Outcome

The clearest evidence of a training effect in Experiment 1 was the increase in both FLIR and visible scores on the posttest. The lowest pretest scores were for the vehicles in the FLIR training; the highest posttest scores were for the same vehicles. Scores increased 49 percentage points, from 32% correct to 81% correct. FLIR scores for those vehicles that were not trained, only tested, increased only 8%. Similar effects were found for visible scores. Scores for vehicles in the training increased 29 percentage points, from 41% correct to 70% correct, compared to a 6% increase for vehicles not trained. Thus there was positive transfer to visible imagery from the thermal imagery.

Another indication of a training effect was the increase in performance from training to the posttest. FLIR scores on the vehicles trained increased from a mean of 66% correct on the first session of the whole-task training set to 81% on the posttest.

Individual differences were evident. For instance, overall, the fastest individual in the Triad condition took 50 min less than the slowest individual. The fastest individual in the Diad condition took 30 min less than the slowest individual. Soldiers' visible and FLIR pretest and posttest scores correlated with each other. So despite the substantial increase in posttest scores on the vehicles trained, those soldiers with the higher initial skills had the higher final skills, and they reached criterion sooner in the training. These findings reinforce the advantages of a self-paced training program that adapts to differences in soldier skill.

Experiment 2: Fixed-Paced with Knowledge of Results versus Self-Paced with Enhanced Knowledge of Results

Experiment 2: Background

Experiment 2 examined the effects of a fixed-paced and a self-paced training strategy and two forms of feedback. These two dimensions, pace and feedback, were not examined independently of each other, however. The fixed-paced training was linked to knowledge of results (KR) feedback, that is, typical feedback on the correctness of responses. This training was executed with the Signature Challenge exercises. The self-paced training was linked to an enhanced form of feedback that directed soldiers to factors critical to their performance. Specifically, when a misidentification was made, the vehicle that was confused with the displayed vehicle was also presented. This allowed the soldier the opportunity to compare the two images and to determine what led to the misidentification. This training was executed with the Vehicle ID exercises.

Signature Challenge is drill-type of training exercise, where soldiers are exposed to a large number of images within a short period of time, relatively quick responses must be made as the exposure time is limited, and immediate feedback on the correctness of each response is provided. There is little opportunity to study each image either before or after responding. Time pressure may in some ways approximate the demands made in combat. Yet the question remains regarding whether this drill format is the best approach to train soldiers when they are first acquiring vehicle identification skills.

Similar issues were raised by Gibson (1947) and his colleagues in their work on aircraft recognition. In the early 1940s, there was debate as to whether aircraft recognition training should consist of flash presentations where individuals learn total forms, or whether it should focus on analyzing the distinctive features of aircraft. At that time, distinctive features were defined as the wings, engine, fuselage, and tail (WEFT). Flash presentations were defined as 1/50th of a second or faster. The rationale for the flash technique was that recognition of aircraft had to be done quickly in combat; the faster, the better. With the flash technique, after each exposure, responses were immediately confirmed or corrected, and the slide shown again to permit individuals to look at the picture while correcting their identification. In an experiment comparing exposures of 1 sec, 1/10th sec, and 1/50th sec, no differences were found on training scores at the end of a 30-hour course that extended over nine weeks. However, test scores were highest when aircraft were displayed for 1 sec, regardless of the prior training condition.

The exposure times used in these experiments by Gibson were much faster than those in Signature Challenge. Yet the general approach was the same. The pace was fixed by the instructor, repeated exposures of vehicles were given, and confirmatory or corrective feedback was provided on each trial.

Conversely, Vehicle ID provided an opportunity to examine or study each image before responding, as no time limit is imposed. In addition to soldiers being told immediately of the correctness or incorrectness of their response, they are also presented with corrective visual feedback when they err. Thus if the vehicle presented is a T72 and they indicate it is an M1, the image of the M1 is then displayed next to the T72 image. This visual feedback gives soldiers the opportunity to compare and contrast the two images. They can determine for themselves why they were confused; what caused them to incorrectly identify the T72 as an M1. If they get the vehicle correct but the aspect wrong, corrective visual feedback showing the correct aspect is provided, and they have the opportunity for self-study before going to the next display.

We viewed this visual corrective feedback as a variation of knowledge of performance. Within the realm of motor skill learning, knowledge of performance refers to information that assists learners in knowing what aspects of their performance contribute to the desired outcome and conversely, what actions need to be modified (Proctor & Dutta, 1995). In some instances, this form of feedback has been shown to be more effective than KR alone. With the Vehicle ID exercises, the visual feedback is tailored to reduce the vehicle confusions unique to each individual.

In a given period of time, soldiers will likely view fewer images in Vehicle ID than in Signature Challenge. There are two reasons for this: the self-pacing feature of Vehicle ID and the requirement to make two responses - vehicle and aspect. Although requiring more time, both features were viewed as having training benefits.

Experiment 2 built on lessons learned from the first experiment. In Experiment 1, we found that although Signature Challenge provided many exposures to vehicle images within a short period of time, soldiers often reached learning plateaus where further practice was not beneficial. They also became frustrated and lost motivation when many repetitions were required to reach criterion. The inexperience of these soldiers with combat vehicle identification may have also been a contributing factor to these learning plateaus. To reduce the frequency of these plateaus, we went from eight to six vehicles, eliminating the ZSU and the M551.

We retained a training set that included all six vehicles. Experiment 1 had shown this training to be essential. In that experiment, soldiers had difficulty when required to identify all vehicles in the training set, despite reaching criterion on the prior part-task sets. When all vehicles were presented, errors occurred both within and across sets. Training on all vehicles was necessary to reduce unexpected confusions among vehicles.

Experiment 2 also examined the benefits of practice in determining aspect angle. The transfer conditions examined transfer from night, white-hot images to both night and day, black-hot images.

It should be noted that, because of computer technology, this experiment differed in a critical respect from previous vehicle and aircraft recognition training research (Gibson, 1947; Haverland & Maxey, 1978; Warnick & Smith, 1989; Whitmore et al., 1968). In these prior efforts, instruction and practice were conducted in a group setting, typically with slide

presentations or scaled models of vehicles or aircraft. With this format, individualized, corrective visual feedback was not possible. With the computer, immediate, corrective, side-by-side visual feedback tailored to each individual's confusions is easily provided. No prior research was found that incorporated training identical to the Vehicle ID format. However in a laboratory setting, Cockrell (1978) investigated a response sensitive training condition where each individual was shown a picture of the correct vehicle after an incorrect identification. This visual feedback was tailored to each individual's errors. But it was sequential visual feedback; it did not coexist with the displayed image as in the Vehicle ID module. Most prior research did, however, incorporate knowledge of results feedback, as in the Signature Challenge module, although the prior research was in group settings where the sequence and pace were controlled by the instructor.

Experiment 2: Purpose

The primary questions addressed were:

- Is a training strategy that requires multiple, timed vehicle identification trials and provides knowledge of results more or less effective than a training strategy, with no time requirement that also provides corrective visual feedback?

- Is transfer greater to the same vehicles displayed in black-hot day imagery or black-hot night imagery, regardless of training strategy?

- Does practice in determining aspect angle improve aspect scores in a transfer condition?

Secondary questions were:

- Does thermal imagery training transfer to visible imagery, regardless of training strategy?

- Does thermal imagery training transfer to vehicles not trained, regardless of training strategy?

- What vehicles tend to be confused with each other during training and testing?

We expected a training strategy that provided corrective visual feedback to be more effective than one that provided only knowledge of results, particularly when the visual feedback was linked with self-pacing. However, because of the configuration of the software in the prototype CVI*Plus* program, part-task training could not be provided in the self-paced condition. The differential effects of part-task training under the two conditions could not be compared. Nor could soldiers in the self-paced condition benefit from part-task training. Consequently, it was possible that positive effects from a self-paced training strategy with visual feedback might not be detected, because of the advantages of part-task training under the fixed-paced condition.

The direct comparison of the Signature Challenge format to the Vehicle ID format resulted in confounding of some training variables. And we were unable to make the necessary changes in the CVI*Plus* software to reduce or eliminate some of the confounds. Nonetheless, we felt that insights regarding training could be ascertained by comparing these two different training strategies.

With regard to transfer to night and day black-hot imagery, Osgood's (1949) transfer surface would predict that transfer would be greater to night, black-hot than to day, black-hot. The similarity between the white-hot night and black-hot night images is greater than the similarity between the white-hot night and black-hot day images. When going from white-hot night to black-hot night, there is a change in one stimulus attribute, that is, the "color" of the hot spots and areas on the vehicle and in the background scene. The hot spots are the same in that they take the same form and are in the same location. One image could be described as a "negative" photo of the other. On the other hand, when going from white-hot night to black-hot day, not only does the hot spot "color" change, but the hot spots themselves can change. For example, hulls and turrets that appear cool and almost invisible at night can appear hot and very distinct during the day. Suspensions that appear hot at night, can appear cooler during the day. Thus the training and transfer stimuli are less similar in this transfer condition because two stimulus attributes change.

No formal expectations were made regarding the effect of training on vehicle aspect. In general, aspect is easier to determine than the vehicle. Therefore, it was questionable whether any substantial improvements would result from training on aspect.

Experiment 2: Method

Participants

Two groups of soldiers participated in Experiment 2. One group was 14 students from the Bradley Leader Course (BLC) at Fort Benning. When the experiment was conducted, these students were halfway through a 16-hour period of instruction on vehicle identification. The second group consisted of 21 soldiers from a Georgia National Guard (NG) Mechanized Infantry unit.

The majority (64%) of the BLC students were Second Lieutenants, while another 21% were First Lieutenants. There was one Sergeant First Class and one Captain. Their ages ranged from 22 to 41 with a mean age of 26.85 years ($SD = 4.9$). Approximately 29% had been in the Army a year or less, 21% had between two and three years, and 36% between four and ten years. The remaining 14% had served eleven or more years. The majority (57%) listed their preceding duty as a student; the remaining had a variety of experience to include prior service as a noncommissioned officer (28%).

Soldiers were asked about their experience with FLIR imagery. Only five (36%) had previous experience with thermal sights. Four of the five had used the thermal sight on the Bradley; one had used the Dragon night tracker.

Of the 21 soldiers from the NG unit, 19% were Privates, 33% were Specialists or Corporals, and 48% held the rank of Sergeant and above. Their mean age was 32.0 years (SD = 9.19) with a range from 19 to 54. Time in the Army was longer for this group than the BLC students. Only 5% had served less than a year in the Army, 19% had between two and three years of service, 19% between four and ten years of service, another 38% between 11 and 20 years, and 24% had over 20 years of Army experience.

Bradley Fighting Vehicle (BFV) gunners constituted 38% of the sample, BFV commanders, 33%, and drivers, 14%. There was also one rifleman and one radio telephone operator. It should be mentioned that 7 (33%) of these soldiers had recently changed from an Engineer military occupational specialty (MOS) to the 11M (Bradley) MOS, so their Infantry experience and particularly their Bradley experience was limited. Overall, half the soldiers had served a year or less in their respective current duty positions, with 28% serving five months or less, and 24% serving six to twelve months. Fourteen percent had been in their position for 13-18 months, another 14% between 19 and 48 months, and 19% had served over 48 months in their duty position.

Sixty-eight percent cited experience with the thermal sight on the BFV. Of this group, two soldiers also had experience with the Dragon night tracker. Sixty-two percent had been to the National Training Center (NTC). Two had participated in Desert Storm.

Design

The design for Experiment 2 is outlined in Table 2-1. It follows the experimental paradigm presented in Table 2 in the Introduction. The sequence of events was pretests, group instruction, individual practice on identifying vehicles, transfer test, and posttests. Soldiers were randomly assigned to either a fixed-paced with KR condition or self-paced with enhanced (visual) KR condition. These two conditions were implemented, respectively, with the Signature Challenge training exercises and the Vehicle ID training exercises. For purposes of this report, the two conditions are labeled "Fixed-KR" and "Self-KR+." All other training and testing events were the same for all soldiers. Soldiers were trained on six vehicles, three main battle tanks and three APCs: M1, T72, M60, M2, LAV, and BTR.

Group instruction. The group instruction on vehicle cues was presented by the military instructor who taught in Experiment 1. An overhead projection system was used to display the Image Library module. Each vehicle was presented with the two-display screen. The left display presented day visible imagery; the right display, white-hot night imagery. Both were presented at the closest range in the Image Library. Thermal cues for each vehicle, using S^4HEET as a framework, were identified as each aspect angle was presented and discussed. The aspects were presented in a the following order: right front oblique, front, left front oblique, right flank, left flank, left rear oblique, rear, and right rear oblique.

After each vehicle was presented, the three-display screen was used to compare vehicles, continuing with the night white-hot images at the closest range. The three tanks were displayed together, followed by the three APCs. For each presentation, vehicle similarities and differences were discussed by aspect angle. Finally, the instructor illustrated changes to thermal signatures as the distance to the vehicle increases. The M1 and T72 were paired on the two-display screen, and shown at each of the four ranges in the Image Library, starting with the closest range and ending with the farthest range, Range 3.

Table 2-1
Design of Experiment 2 and Sequence of Events

<table>
<tr><th>Sequence of Events</th><th colspan="2">Description</th></tr>
<tr><td>1. Pretests</td><td colspan="2">FLIR test, then Visible test. Same pretest as Experiment 1.</td></tr>
<tr><td>2. Group Instruction on Vehicles and Vehicle Cues</td><td colspan="2">Overhead projection using Image Library. White-hot night at Range 2.
-Vehicles were: M1, M60, T72, M2, LAV, and BTR.
-Each vehicle presented at closest range. Side-by-side comparisons with visible image on 2-display screen. Thermal cues cited for each aspect.
-Compared night images at closest range on the 3-display screen.
M1, T72, and M60 (tanks)
M2, LAV, and BTR (APCs)
-Showed changes in imagery with distance:
Paired M1 and T72, at close-up range, Ranges 1, 2, and 3.</td></tr>
<tr><td>3. FLIR ID Training</td><td colspan="2">Random assignment of soldiers to Fixed-KR and Self-KR+ conditions
-Showed all aspects of each vehicle at Range 2 in night, white-hot.
-Pass criterion was no more than 4 errors.</td></tr>
<tr><td></td><td>Fixed-paced with KR
(SC exercises)
(n = 18)

① M1, M60, T72 (30 images, 10 per vehicle)
② M2, LAV, BTR (30 images, 10 per vehicle)
③ All 6 vehicles (48 images, 8 per vehicle)</td><td>Self-Paced with KR+
(Vehicle ID exercises)
(n = 16[a])

All 6 vehicles (48 images, 8 per vehicle). Same as last vehicle set in SC exercises.</td></tr>
<tr><td>4. Transfer test (see Table 2-2)</td><td colspan="2">Scored Vehicle ID test module used to examine transfer to night and day black-hot thermal images.</td></tr>
<tr><td>5. Posttests</td><td colspan="2">Same as Pretest.</td></tr>
</table>

[a] Original n was 17, but one individual withdrew.

The group instruction was given to all BLC participants at the same time. It was repeated three times to accommodate the NG participants. Each group presentation took about 45 min.

Training. The effectiveness of a fixed-paced presentation of images with typical knowledge of results feedback on each trial was compared to a self-paced presentation of images with visual, corrective knowledge of performance feedback on each trial. Ideally, everything else, such as the training sequence, should have been the same. However, this degree of experimental control was not possible.

Fixed-paced training could only be executed with the Signature Challenge module, where the experimenter or instructor could control exposure time of the vehicles. Only confirmatory and corrective feedback is provided. If correct, the vehicle response button turns green. If incorrect, the response button turns red, and the correct response button turns green. The fixed, rather short exposure time, is a distinctive feature of the Signature Challenge module. In addition, there is very little delay after the soldier responds and receives feedback to the presentation of the next image. Once vehicle selection is made, the soldier cannot change it.

Self-paced training could only be executed with the Vehicle ID module. By default, trials are not timed in this module, as the instructor/experimenter has no means of adjusting the vehicle display time. Vehicle ID provides KR via green and red buttons as is done with Signature Challenge. However, it also provides corrective visual feedback, that being its most distinctive training feature. In contrast with SC, a soldier can change vehicle and aspect angle selections, if desired, before feedback is provided. And in case of errors, the two vehicles and/or aspects can be studied and compared before progressing to the next vehicle.

The automated gates and capability for multiple, sequential vehicles sets in Signature Challenge, but not in Vehicle ID, made it difficult to control for all other factors. In Signature Challenge the experimenter/instructor pre-sets the pass criterion for a vehicle set. If the soldier does not meet this criterion, the program automatically presents the same vehicle set again. This cycle is repeated until the criterion is met. Additional vehicle sets can follow the first set, with soldiers required to meet criterion on each set in the sequence. Vehicle ID did not have these two features.

Consequently, some compromises were made with the design. The results of Experiment 1 indicated that presenting all six vehicles in a single Signature Challenge set would be too difficult for most soldiers. Therefore, we applied a pure-part training strategy in the Fixed-KR condition. Soldiers trained first on all the tanks, then all the APCs, and finally all six vehicles. The last set contained all vehicles, as Experiment 1 showed the need for this type of training. This last set also provided a comparison with the Self-KR+ condition using the Vehicle ID module, which permitted only one training set.

In Vehicle ID, we had to train soldiers on all six vehicles simultaneously. However, we thought the self-paced nature of Vehicle ID would reduce the likelihood of the learning plateaus. We did, however, impose a gate controlled by the research staff. Thus each soldier repeated the exercise until the criterion was achieved.

All training imagery was white-hot night at Range 2. Range 2 was selected because we expected the soldiers in this experiment to have more experience with vehicle identification and

thermal sights than those in Experiment 1. In the Fixed-KR training, the vehicle aspects presented in the first two vehicle sets (tanks then APCs, see Table 2-1) were the same as those in Experiment 1, but at Range 2 instead of Range 1. In the last Fixed-KR vehicle set and in the single Self-KR+ set, only the eight aspects of each vehicle at Range 2 were presented. The criterion for passing a vehicle set was no more than four errors (85% correct with 30 displays and 91% correct with 64 displays. In summary, the two training conditions were not equivalent in terms of the sequence of vehicle sets and the number of vehicle displays. If a soldier met criterion on the first session, he would have been exposed to 108 images in Fixed-KR and to 48 images in Self-KR+. In Fixed-KR, soldiers had seven sec to respond to the vehicle sets with 30 displays, and ten sec to respond to the vehicle set with 48 displays.

Seven computers were available. To accommodate all participants, two individual training sessions were held for the BLC; three training sessions for the NG.

Testing for transfer. The Vehicle ID test module was used for the transfer test. All the oblique angles plus the two flanks assessed transfer to black-hot images at day and night. The front and rear aspects were the no transfer conditions. Range 2 was used for all images.

Some counterbalancing of the imagery was required to keep the test length reasonable. For the transfer conditions, every vehicle was presented in black-hot, both day and night, as a front oblique, rear oblique, and flank. However, for the night, black-hot condition, three of the six vehicles were displayed from the right view and half from the left view. For the day, black-hot condition, the views were switched. The test had 48 displays: 36 transfer images and 12 no transfer images. Table 2-2 outlines the test format.

Table 2-2
Experiment 2: Format of Transfer Test

Imagery	Transfer						No Transfer	
	Front Oblique		Rear Oblique		Flank			
	Right	Left	Right	Left	Right	Left	Front	Rear
Night Black-Hot	BTR M2 T72	M1 LAV M60	M1 LAV M60	BTR M2 T72	BTR M2 T72	M1 LAV M60		
Day Black-Hot	M1 LAV M60	BTR M2 T72	BTR M2 T72	M1 LAV M60	M1 LAV M60	BTR M2 T72		
Night White-Hot							All vehicles	All vehicles

Pretests and posttests. The pretests and posttests were identical to those in Experiment 1. As before, visible and FLIR versions of each were given. As noted previously, only six of the

ten vehicles on the pre- and posttests were trained in this experiment: M1, T72, M60, M2, LAV and BTR. Training did not cover the BMP, M113, HMMWV, and ZSU.

Measures of Performance

The pretest, posttest, transfer test, and training conditions required soldiers to identify the vehicle displayed. Vehicle images were presented randomly within all exercises and tests. For the pre- and posttests, soldiers identified the vehicle by selecting a name from a list of all the 14 vehicles in CVI*Plus*. For the Fixed-KR training, only the vehicles being displayed were possible responses. Depending on the vehicle set, this involved a choice of three or six vehicles. For the Self-KR+ training and the transfer test using Vehicle ID, the response buttons for all 14 vehicles in CVI*Plus* were automatically displayed as options. Soldiers were told that only the six vehicles being trained would be presented. They were given a list of these vehicles in order to make the Vehicle ID module response format similar to Signature Challenge.

The Vehicle ID exercises and tests required soldiers to identify the aspect angle as well. The other criterion measure was response time. Response times on each trial in each exercise and test were recorded. Total time to complete training was tabulated. The number of sessions required to reach criterion during training was recorded as well.

Experiment 2: Results

Program Completion

The BLC students had two periods of approximately 1.5 hr each for training and testing in addition to the group instruction time. Two students did not have time to meet the 91% training criterion on the Self-KR+ vehicle set, but they did take all tests. With the NG, the time constraints were more severe. Resolving computer problems also consumed some of the training time. Over half the NG soldiers were allowed to take the tests, even though they had not achieved criterion on the final vehicle set in their respective training condition. As with Experiment 1, many soldiers were very close to passing the final, all-vehicle set. Given the time constraints and the computer problems, we changed the criterion on this set from 91% to 85% (no more than 7 errors). Table 2-3 gives the status of soldiers at each phase of the experiment.

Training and Testing Time

The times for each major training and testing event are in Figure 2-1. The times include the pretests, the posttests, group instruction, SC and Veh ID practical exercises, and transfer test. Administrative and computer down times are excluded. Mean times for the Fixed-KR and Self-KR+ conditions are displayed, as well as the times for the fastest and slowest individuals in each condition. Although everyone was not able to train to the same criterion, the fastest and slowest individuals within each condition did reach the 91% criterion. The mean times include all participants, regardless of their status at the end of training. Group instruction was 45 min.

Table 2-3

Experiment 2: Number of Soldiers Completing Each Training Exercise and Test

Status on Training Exercises and Tests	Fixed-KR	Self-KR+	Total
Completed Pretests	18	16	34
Met original 91% criterion			
1st Vehicle Set	18	NA	---
2nd Vehicle Set	17[a]	NA	---
All-Vehicle Set	13[b]	6	19
Met revised 85% criterion for All-Vehicle Set	14[b]	10	24
Completed Vehicle ID Transfer Test	17[c]	16	33
Completed Posttests	18[d]	16	33
Soldiers with Complete Data	16	16	32

[a] One soldier did not take this vehicle set

[b] 16 soldiers available. No data on two soldiers as one did not take this vehicle set and data from another was lost due to a computer problem.

[c] One soldier's data was lost because of a computer problem.

[d] Visible posttest missing on one soldier.

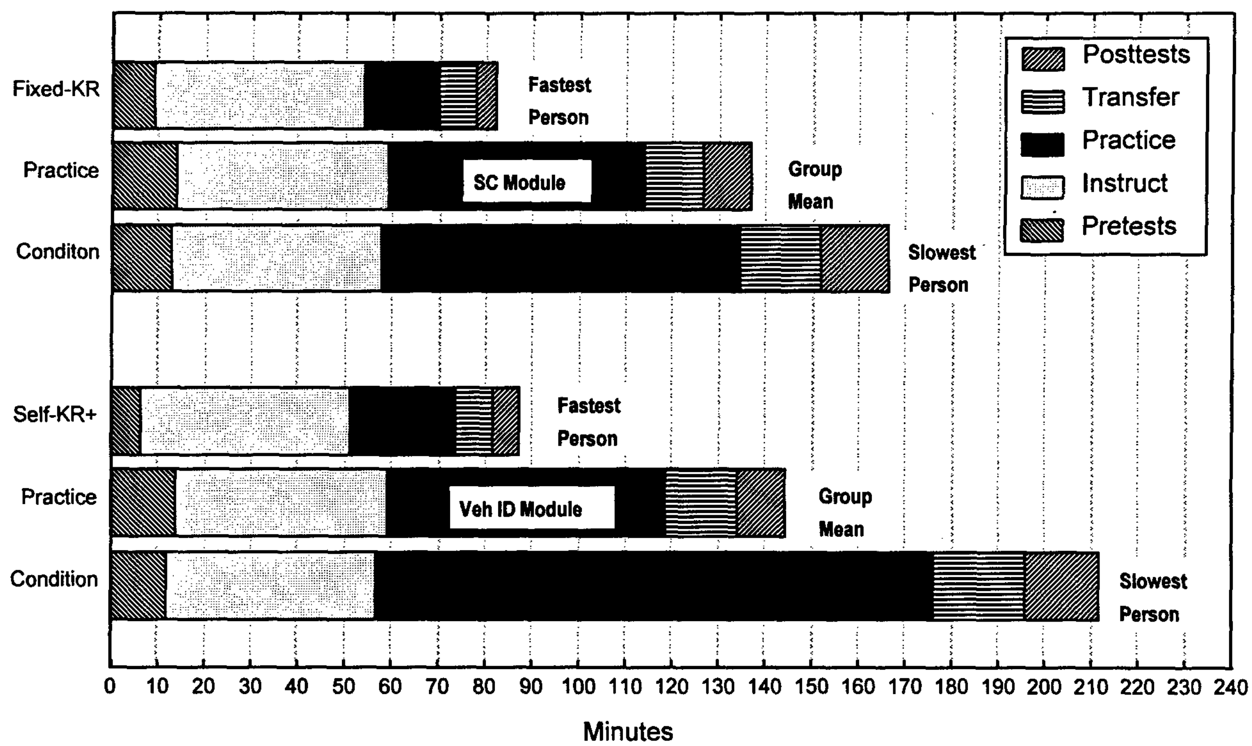

Figure 2-1. Experiment 2: Training and testing times for the Fixed-KR and Self-KR+ conditions (group means and fastest and slowest individuals).

The time profiles for the two conditions and for the fastest individuals within each condition were quite similar. However, the slowest person in Self-KR+ required 45 minutes longer than the slowest person in Fixed-KR. The longer times for Self-KR+ reflect the non-timed feature of Veh ID. If there had been no time constraints during the experiment, the Self-KR+ mean time would have been even greater, as all soldiers would have had the opportunity to reach criterion. Table C-1 provides a further breakdown of the times.

Training Exercises

Although the soldiers from the BLC and NG differed in terms of Army experience and rank, the data from both groups were combined for the statistical analyses. All soldiers in the Fixed-KR condition met criterion on the first two vehicle sets. As cited previously, computer problems and time constraints impacted soldier ability to meet criterion when all six vehicles were presented in Fixed-KR and in Self-KR+. The criterion was revised from 91% to 85%. With the 91% criterion; 81% of those in Fixed-KR passed compared to 38% in Self-KR+. With the 85% criterion, 87% of those in Fixed-KR passed compared to 63% in Self-KR+. There was a significant difference in the proportion of soldiers who passed in Fixed-KR and Self-KR+ when the 91% criterion was used, $\chi^2(1) = 4.14$, $p < .0418$, but not when the 85% criterion was used, $\chi^2(1) = .95, p < .3291$.

Eight soldiers, two in Fixed-KR and six in Self-KR+, did not meet the 85% criterion. Performance on the last session of the all-vehicle set was examined. For the two soldiers in Fixed-KR, their scores were 83% and 77%, for a mean of 80%. These soldiers stopped training on the all-vehicle set after one and three sessions respectively. For the six soldiers in Self-KR+, their mean score was 60%. The lowest score was 46%; the highest 81%. Four of the six completed two sessions on the all-vehicle set, one completed four, and one completed only one.

Both the 91% and 85% criteria were used to examine the <u>number of sessions</u> needed to pass the <u>all-vehicle set</u>. These findings indicate the amount of additional effort required to train to a very high standard (see Table 2-4). The 91% criterion required many sessions for some soldiers, but the same soldiers reached the 85% criterion in fewer sessions. Of note is one soldier in Fixed-KR who required 19 sessions to meet 91%, but who had achieved the 85% criterion on the third session. Obviously, for this individual, this form of repeated and quick exposure to the images, without some additional help or remediation, was unproductive.

Soldiers in the Fixed-KR condition received part-task training on the vehicles. Consequently, they were expected to do better on the first session of the all-vehicle set than soldiers in the Self-KR+ condition, whose first session on the all-vehicle set was their initial exposure to the vehicles. Fixed-KR first-session scores on the all-vehicle set (M = 82%, SD = 9) were, in fact, significantly higher than Self-KR+ first-session scores (M = 68%, SD = 15), $F(1, 30) = 9.56$, $p < .0042$. Of interest is that when only those soldiers that met the 85% criterion on this set were compared, there were no significant differences. The mean score for those in Fixed-KR stayed the same, but the mean score for those in Self-KR+ increased to 76% correct.

Table 2-4
Experiment 2: Number of Sessions to Reach Criterion During Training on the All-Vehicle Set

	Fixed-KR			Self-KR+		
	# Sessions		# Soldiers[a]	# Sessions		# Soldiers[a]
Criterion	*M* (*SD*)	Range		*M* (*SD*)	Range	
91%	4.76 (4.93)	1 - 19	13	2.83 (0.98)	2 - 4	6
85% (soldiers who met 91% only)	2.15 (1.57)	1 - 6	13	2.00 (0.89)	1 - 3	6
85%	2.07 (1.54)	1 - 6	14	2.60 (1.77)	1 - 7	10

[a] Number of soldiers reaching criterion from a total of 16 in each training condition.

Table 2-5 presents the mean identification scores per vehicle on the first session of each vehicle set for all soldiers. For Fixed-KR training, scores on each vehicle increased from the part-task sets to the whole-task set, except for the M2, which had the highest initial score. Initial scores for each vehicle on the first session of the Self-KR+ set were lower than corresponding vehicle scores on the first session of the Fixed-KR all-vehicle set.

Table 2-5
Experiment 2: Mean Vehicle Identification Scores (% correct) on the First Session of Each Vehicle Set

	Training Condition			
	Fixed-KR Vehicle Sets			Self-KR+ Vehicle Set
Vehicle	Tanks (*n* = 18)	APCs (*n* = 17)	All Vehicles (*n* = 16)	All Vehicles (*n* = 16)
M60	72		87	81
M1	59		80	55
T72	59		73	68
M2		92	92	75
BTR		72	83	67
LAV		62	75	61
All Vehicles	60	75	82	68

Note. There were 10 trials per vehicle per individual in the Fixed-KR tank and APC sets; 8 trials per vehicle per individual in the all-vehicle sets.

Not all soldiers in the Fixed-KR condition mastered each part-task vehicle set on the first session. On the tank set, the number of sessions ranged from 1 to 12 with an average of 6 sessions. Half the soldiers met criterion in 4 sessions or less; half met criterion in 6 to 12 sessions. On the APC set, the number of sessions ranged from 1 to 11, with an average of 4 sessions. Half the soldiers met criterion in 1 or 2 sessions; 35% in 4 to 6 sessions, and the

remaining 18% in 10 or 11 sessions. Overall, initially soldiers knew the APC set better than the tank set (see Table 2-5), and consequently required fewer sessions to achieve criterion.

A major difference in the two training conditions was the response time for each trial. Those in the Fixed-KR condition had a time limit; seven sec on each trial for the part-task vehicle sets and ten sec for the all-vehicle set. Those in the Self-KR+ condition were not timed, and had to select both the vehicle and the aspect. Descriptive data on these times are in Table 2-6. Soldiers in the Fixed-KR condition did not use the maximum times allowed, even though the instructor reminded them that they were responding within 2 or 3 sec and could easily study each image longer. Clearly, the soldiers in Self-KR+ took advantage of the non-timed trials and studied the images before responding. The requirement to make two responses in Self-KR+, as opposed to one in Fixed-KR, cannot account for the fact that soldiers in Self-KR+ responded 10 times slower than those in Fixed-KR. The findings for the Self-KR+ group clearly indicate individual differences, with the slowest soldier taking a little over a minute to respond to each display; the fastest, about 15 sec.

Table 2-6
Experiment 2: Distribution of Soldiers' Mean Response Times to Each Trial on the First Session of the Training Sets

	Training Condition				
	Fixed-KR			Self-KR+	
	Tanks (n = 18)	APCs (n = 17)	All Vehicles (n = 16)	All Vehicles (n = 16)	
Mean	% Soldiers			Mean	% Soldiers
< 2 sec	17%	12%	25%	< 20 sec	44%
2-3 sec	29%	70%	44%	20-30 sec	19%
3-4 sec	33%	8%	31%	40-50 sec	19%
>4 sec	22%	0%	0%	>50 sec	19%
	Mean (sec)				Mean (sec)
Fastest Person	1.74	1.66	1.79	Fastest Person	14.37
Slowest Person	5.18	3.44	3.48	Slowest Person	63.96
Group Mean	3.16	2.56	2.65	Group Mean	28.33

Note. There were 30 trials per individual for the tank and APC sets in Fixed-KR; 48 trials for the all-vehicle sets in Fixed-KR and Self-KR+.

Response time was also examined as a function of the vehicle displayed. Both groups responded fastest to the M60, 18.30 sec for Self-KR+ and 2.19 sec for Fixed-KR. Additional information on response times in the all-vehicle set is in Table C-2.

A more detailed examination of the training data was made to determine which, if either, training strategy was better for soldiers who did not meet the 91% criterion quickly. This analysis was restricted to soldiers who had more than three sessions and who achieved either the 91% or 85% criterion. Five soldiers in each training condition fit this definition. Even though

some who reached the 91% criterion had achieved 85% on a prior session, all their sessions on the all-vehicle set in Fixed-KR and Self-KR+ were included in the analysis. Table C-3 summarizes the initial scores for these ten individuals, number of sessions required to reach criterion, and response times. Learning curves for these individuals are in Figures C-1 and C-2.

Given the limited sample size, no inferential statistics were computed. However, several trends are evident. First, those in Fixed-KR required more sessions to reach criterion than those in Self-KR+, yet they had the higher pretest FLIR and first-session scores. Second, the learning curves show that four of the five soldiers in Self-KR+ training showed a steady increase in progression toward criterion, versus two of the five soldiers in Fixed-KR training. It should be noted that the four soldiers who took the "uneven road" to criterion often had fast response times within their respective training conditions.

In general, Self-KR+ seemed to provide the better learning environment for soldiers who did not reach criterion quickly. The non-timed trials in Self-KR+ combined with the opportunity to change a response and the corrective visual feedback probably enhanced the learning process. The Fixed-KR condition may have become more like a "test" to the soldiers. Quick and unintentional erroneous selections with the computer mouse could have also accounted for some difficulties encountered by those in Fixed-KR. Finally, these findings point to a need for remedial training when soldiers have not reached criterion within about three sessions. They need to exit the training exercises, study the vehicles with which they are having difficulties, and then return to the training exercises.

Which vehicles did soldiers confuse with each other? Did similar vehicle confusion patterns occur regardless of the training condition? To address these questions, the first session of each vehicle set was examined. In both Fixed-KR and Self-KR+ training, the tanks tended to be confused with each other, the BTR and LAV were confused with each other, and the LAV also tended to be misidentified as the M2 (Table 2-7). After reaching criterion on the part-task sets, these confusions were reduced for soldiers in Fixed-KR on the all-vehicle set. The confusions that remained, those greater than 10%, were between the M1 and the T72, and the LAV being misidentified as the BTR.

The main difference between the two conditions on the all-vehicle set was that soldiers in Fixed-KR had fewer confusions between the two classes of vehicles - tanks and APCs. Confusions between these two classes averaged 4% per vehicle on the first-session of the all-vehicle set for Fixed-KR, reflecting the positive effects from the prior part-task training. In contrast, for Self-KR+, confusions between the two classes of vehicles averaged 10% per vehicle.

The Self-KR+ training exercise also required soldiers to determine aspect angle. As shown in Table C-4, the two flanks were the easiest to identify; then the rear and rear oblique aspects, followed by the front and the front oblique aspects. The front and rear views were confused with each other.

Table 2-7

Experiment 2: Vehicle Confusion Matrices from the First Session of Each Vehicle Set in Fixed-KR and Self-KR+ Training Exercises

Vehicle Class	Vehicle Displayed and Training Condition	Vehicle Response					
		Tank			APC		
		M1	M60	T72	M2	BTR	LAV
Tank	M1 Fixed-KR	**59**	15	25	----	----	----
	Fixed-KR	**80**	5	12	3	0	1
	Self-KR+	**55**	11	20	9	1	4
	M60 Fixed-KR	10	**72**	18	----	----	----
	Fixed-KR	4	**88**	5	3	0	1
	Self-KR+	4	**81**	8	3	2	2
	T72 Fixed-KR	26	15	**59**	----	----	----
	Fixed-KR	19	3	**73**	2	2	1
	Self-KR+	12	11	**68**	5	1	4
APC	M2 Fixed-KR	----	----	----	**92**	4	4
	Fixed-KR	1	2	1	**92**	3	1
	Self-KR+	5	2	9	**75**	5	5
	BTR Fixed-KR	----	----	----	10	**72**	18
	Fixed-KR	3	1	1	4	**83**	8
	Self-KR+	5	3	7	6	**67**	11
	LAV Fixed-KR	----	----	----	13	25	**62**
	Fixed-KR	0	4	2	5	14	**75**
	Self-KR+	0	4	5	12	16	**61**

Note. Percent correct responses are in bold on the diagonal. Confusions are on the off-diagonal. The first entry in each cell refers to either the tank or APC vehicle set in Fixed-KR. The second entry is the all-vehicle set in Fixed-KR. The last entry is the all-vehicle set in Self-KR+. Percentages for first row of Fixed-KR are based on 170 trials; number of trials for second row of Fixed KR and Self-KR+ is 128.

Transfer of Skill

The transfer data were examined from two perspectives: all participants and only those meeting the 85% training criterion. The latter analysis provided a more comparable picture of the effects of the two training conditions, as soldiers had obtained a similar level of proficiency in both groups. Differences between the two sets of findings are noted. From this point on, these two groups are referred to as the entire sample and the Pass85 sample.

First, <u>identification scores</u> on the no transfer conditions were examined. The no transfer imagery was the front and rear aspects of all six vehicles presented in white-hot night imagery at Range 2 (refer to Table 2-2). There were no significant differences between Fixed-KR and Self-KR+ on these scores, with all soldiers as well as with those who met the 85% criterion. Of interest, however, was that the difference between identification scores for the training conditions was substantially greater when based on data from all soldiers. For all soldiers, the Fixed-KR

score was 85% (*SD* = 20); the Self-KR+ score, 70% (*SD* = 22). For those who reached criterion, the Fixed-KR score was 86% (*SD* = 11); the Self-KR+ score was 84% (*SD* = 12).

Next, identification scores on the transfer conditions were examined. These were the flanks, front obliques, and rear obliques presented at Range 2 in black-hot, night and day images. A 2 x 2 x 3 (training condition x day-night x aspect) ANOVA with repeated measures on the last two factors was conducted. Due to the counterbalancing in the design of the transfer test, the three levels of the aspect factor were both flanks, both front obliques, and both rear obliques. The analyses on the two samples produced slightly different results. In both cases, black-hot night scores were higher than black-hot day scores, and identification scores were highest for the flanks, next highest for rear obliques, and lowest for front obliques. When only the soldiers who reached the 85% criterion were examined, those in the Self-KR+ training condition scored significantly higher overall. These results are detailed in Table 2-8. The findings also show that the scores for those who reached the 85% criterion were higher and tended to be less variable than the entire sample.

Table 2-8
Experiment 2: Mean Identification Scores (% correct) on Transfer Imagery for Significant Effects

Comparison	Soldier Sample			
	Entire Sample ($n = 33$)		Pass85 Sample ($n = 24$)	
	M	*SD*	*M*	*SD*
Training				
Self-KR+	78	19	88[a]	13
Fixed-KR	74	18	79[a]	9
Time of Day				
Night Black-Hot	80[b]	16	88[c]	14
Day Black-Hot	71[b]	20	79[c]	10
Aspect				
Flanks	81[d]	19	88[e]	14
Rear Obliques	76[d]	21	84[e]	18
Front Obliques	69[d]	23	78[e]	17

[a] Training Condition: $F(1, 22) = 5.99, p < .0227$
[b] Night-Day: $F(1, 31) = 14.69, p < .0005$
[c] Night-Day: $F(1, 22) = 9.07, p < .0064$
[d] Aspect: $F(2, 62) = 16.14, p < .0000$
[e] Aspect: $F(2, 44) = 9.22, p < .0004$

A follow-on analysis was conducted with the Pass85 sample that included the two training conditions and the three types of imagery (white-hot night, black-hot night, and black-hot day); a 2 x 3 ANOVA with repeated measures on the last factor. The interaction between these two factors, $F(2, 44) = 72.93$, $p < .0483$, showed the differential impact of transfer on soldiers in the two training conditions. On the no transfer images, scores for the two training

conditions were identical. But for both of the black-hot transfer conditions, Fixed-KR scores were 9 percentage points lower than Self-KR+ scores. For both training conditions, black-hot night scores were higher by 9 percentage points than black-hot day scores.

In summary, for the Pass85 sample, there was no difference on vehicle identification scores between the two training conditions on the no transfer imagery (the front and rear aspects, white-hot night). But on the transfer imagery (the other six aspects, black-hot, day and night), scores for the soldiers in the Self-KR+ condition were higher. Vehicles were easier to identify with the black-hot night imagery than the white-hot day imagery. Vehicles were easier to identify when the flank was displayed.

Vehicle confusions for the entire sample and the 85% criterion sample are in Table C-5. For each vehicle, more confusions occurred for the entire sample than for those who met the 85% training criterion. For the Pass85 sample, a primary difference between the two training groups was that the confusions between the M1 and T72 remained relatively high for Fixed-KR, but decreased for Self-KR+. In addition, the Self-KR+ group was less likely to misidentify the LAV as a BTR.

The Vehicle ID test also required soldiers to indicate the aspect of each vehicle. The analysis of variance procedures applied to the identification scores were also applied to the aspect angle responses. For both samples, aspect scores were significantly higher for those trained with Self-KR+ than with Fixed-KR. In addition, flanks were easier to determine than the obliques, with the front oblique the hardest (Table 2-9). Of interest was that the aspect angle scores paralleled the identification scores; when it was hard to identify a vehicle from a particular aspect, it was also hard to determine the aspect itself.

Table 2-9

Experiment 2: Mean Aspect Scores (% correct) on Transfer Imagery for Significant Effects

Comparison	Soldier Sample			
	Entire Sample (n = 33)		Pass85 Sample (n = 24)	
	M	*SD*	*M*	*SD*
Training				
Self-KR+	78[a]	12	84[b]	11
Fixed-KR	67[a]	14	71[b]	8
Aspect				
Flanks	86[c]	16	88[d]	14
Rear Obliques	73[c]	20	79[d]	17
Front Obliques	59[c]	19	67[d]	17

[a] Training Condition: $F(1, 31) = 5.37, p < .0261$

[b] Training Condition: $F(1, 22) = 12.00, p < .0022$

[c] Aspect: $F(2, 62) = 26.15, p < .0000$

[d] Aspect: $F(2, 44) = 14.16, p < .0000$

Table C-6 summarizes aspect angle confusions. Similar results occurred for the entire sample and the Pass85 sample. Clearly, front and rear aspects were confused with each other, even though they were the no transfer conditions. For the aspects presented with the transfer imagery, the significantly lower overall score for the Fixed-KR condition reflects, to some extent, confusions created by two rear obliques and the right front oblique which did not occur for Self-KR+.

An analysis of the aspect scores was conducted with the Pass85 sample to determine if these scores were affected by the training and transfer conditions. This was a 2 x 3 ANOVA that paralleled the ANOVA conducted on the identification scores. There was a significant main effect for training, $F(1, 22) = 12.17, p < .0020$. Self-KR+ aspect scores averaged 85% compared to a Fixed-KR aspect score of 73%. This effect occurred even though aspect scores were not used as the criterion for exiting the Vehicle ID training exercise. Although there was no significant interaction between the two factors, it appeared that training soldiers on determining vehicle aspect in the Vehicle ID training exercises helped most with new imagery. Aspect scores for Self-KR+ were the same, 85%, for each training condition. However, for Fixed-KP, aspect scores for the black-hot night images (no transfer) were the highest (78%), with lower aspect scores for black-hot day (73%) and black-hot night (70%).

Response times were examined for both training conditions as a function of vehicle and aspect presented (see Tables C-7 and C-8). For the entire sample, the Fixed-KR group's responses were 3 sec faster than the Self-KR+ group (19.52 vs 16.82 sec). For both groups (entire sample), soldiers responded fastest to the M60, within 13 and 15 sec. With regard to aspect, fastest times were for both flanks, also within 13 to 15 sec. For the Pass85 sample, the main trend was for response time by the Self-KR+ group to become faster, decreasing overall by 5.5 sec. For both samples, considerable individual variability in response times to each trial occurred, but the Fixed-KR group was less variable.

Pretests and Posttests

Pre- and posttest vehicle identification scores were analyzed separately for the entire sample and for those who met the 85% training criterion. A 2 x 2 x 2 x 2 (training condition x pre-post x visible-FLIR x train-no train) ANOVA with repeated measures on the last three factors was conducted. For both analyses, there were main effects for pre-post and the visible-FLIR factors, and a significant two-way interaction between pre-post and train-no train. For the Pass85 sample, there was also a train-no train effect and a three-way interaction among training condition, pre-post and visible-FLIR. Scores for the entire sample are in Tables C-9 and C-10. Results for the Pass85 sample are in Table 2-10 and Table C-10.

As in Experiment 1, both the FLIR and visible scores for the vehicles in the training program increased significantly, while scores for vehicles not in the training did not change. For the Pass85 sample, the average increase for the vehicles in the training was 20 percentage points (see Figure 2-2). Similar results were found for the entire sample (Figure C-3).

Table 2-10

Experiment 2: Mean Percentage of Vehicles Identified Correctly on the Pre- and Posttests as a Function of FLIR Training and Type of Imagery for the Pass85 Sample

Vehicle	Visible Imagery			FLIR Imagery		
	Pretest	Posttest	Difference: Post-Pre	Pretest	Posttest	Difference Post-Pre
Vehicles with FLIR Training						
M60A3	72	98	26	69	97	28
M2/M3	90	96	6	83	94	11
M1	92	93	1	76	89	13
BTR-80	53	80	27	62	94	32
T72	78	94	16	67	81	14
LAV-25	62	87	25	46	74	28
Vehicles without FLIR Training						
HMMWV	93	94	1	92	96	4
M113	93	96	3	94	94	0
BMP-2	62	59	-3	46	46	0
ZSU 23-4	51	43	-8	36	36	0
Trained	74	91	17	67	88	21
Not Trained	75	73	-1	67	68	1
Overall[ab]	75	84	9	67	80	13

Note. Vehicle means based on 3 trials (displays) per soldier; 24 soldiers.
Significant main effects for Pre-Post, Visible-FLIR, and Trn-NoTrn factors.
Pre-Post: $F(1, 21) = 19.97$, $p < .0002$
Visible-FLIR: $F(1, 21) = 12.57, p < .0019$
Trn-NoTrn: $F(1, 21) = 8.56, p < .0080$
Significant three-way interaction - Trn Cond x Pre-Post x Vis-FLIR: $F(1, 21) = 5.15, p < .0337$
[a] Pre-Post x Trn-NoTrn: $F(1, 21) = 16.45, p < .0006$
[b] Vehicle means weighed equally in overall mean.

For the Pass85 sample, the significant three-way interaction with the two training conditions showed that scores for the soldiers in the Self-KR+ group increased more from pretest to posttest on all FLIR imagery than did scores for soldiers in the Fixed-KR group (18% versus 5%). Both groups increased by the same amount on visible imagery (see Figure 2-3).

Vehicle confusion matrixes for the entire sample and the Pass85 sample are in Tables C-11 and C-12, respectively. As expected, fewer confusions existed on the posttests. Of the vehicles in the training, the main confusions (greater than 10%) that remained were the LAV being misidentified as either the M2 or the BTR; the T72 as a M1 (for the entire sample only); and the BTR as a LAV. However, of the vehicles not trained, pre- and posttest confusions were

relatively similar, particularly for the BMP and the ZSU. Both were typically confused with many of the other tracked vehicles.

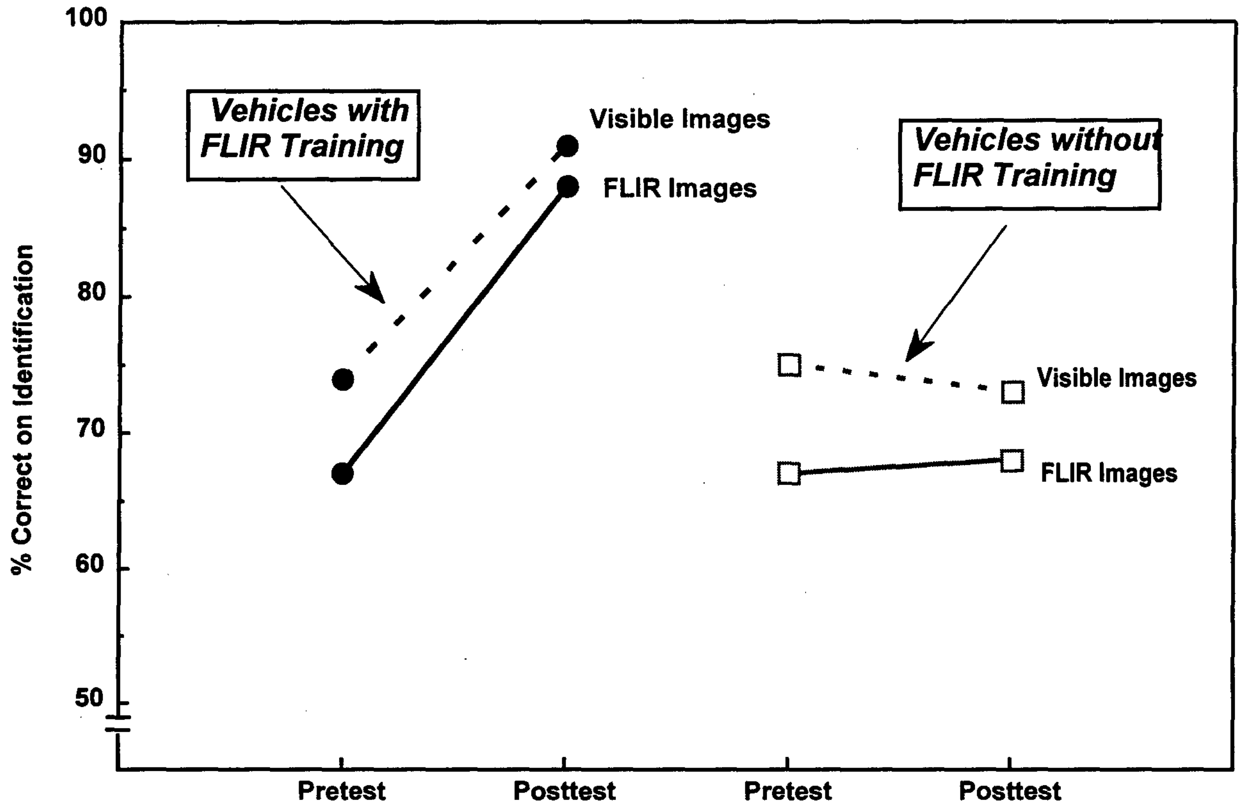

Figure 2-2. Experiment 2. Change in identification scores for the Pass85 sample from pretest to posttest for FLIR and visible images as a function of FLIR training.

Response times on the pretests and posttests were also examined. A 2 x 2 x 2 x 2 (training condition x pre-post x visible-FLIR x train-no train) ANOVA with repeated measures on the last three factors was conducted for the entire sample and the Pass85 sample. Significant effects were the same for both analyses. (See Table C-13 for means and F values for the entire sample and Pass85 sample.) There were main effects for pre-post, visible-FLIR, and train-no train. As in Experiment 1, there was a significant two-way interaction between pre-post and visible-FLIR. For the Pass 85 sample, FLIR times were about twice as fast on the posttest as on the pretest, decreasing from a mean of 14.88 sec (*SD* = 6.43) to a mean of 8.60 sec (*SD* = 3.60), and approached the reaction times to the visible imagery. On the other hand, visible times remained the same, averaging 7.45 sec (*SD* = 2.71) on the pretest and 6.33 sec (*SD* = 3.58) on the posttest. Lastly, there was a significant three-way interaction with training condition, pre-post, and train-no train. In essence, for the vehicles in the training, response times dropped about 4 sec in each training condition. But for the vehicles not in the training, times dropped 5.5 sec for the Fixed-KR condition, compared to 2 sec for the Self-KR+ condition (Pass85 sample).

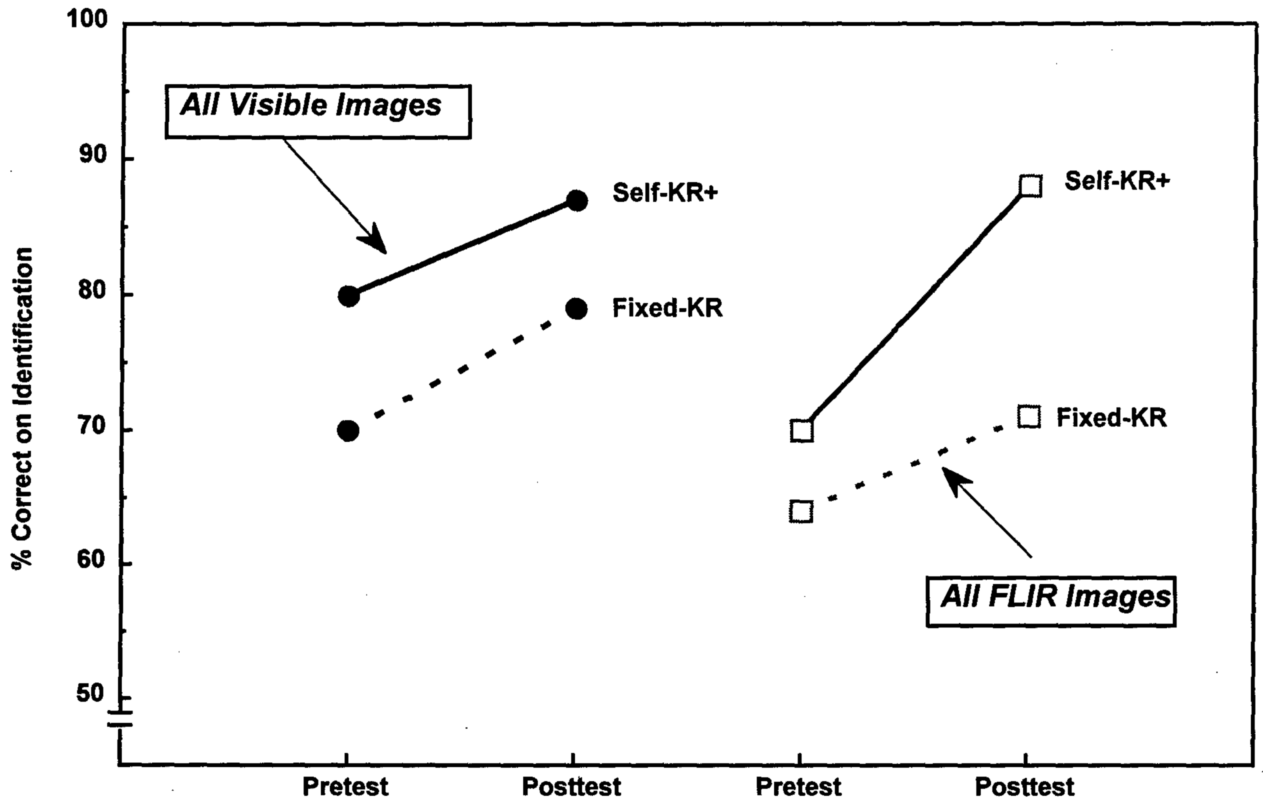

Figure 2-3. Experiment 2. Change in identification scores for the Pass85 sample from pretest to posttest for FLIR and visible images as a function of training condition.

Correlations Among Scores

Correlations among the test and training scores were computed. Pairwise deletion of cases was used when incomplete data occurred. Table 2-11 presents the correlations among the pretest and posttest scores. As in Experiment 1, these correlations were high.

Table 2-11
Experiment 2: Correlations Between Pre- and Posttest Scores

	Pre Visible	Pre FLIR	Post Visible	Post FLIR
Pre Visible	....	.93***	.90***	.88***
Pre FLIR		....	.86***	.84***
Post Visible			....	.85***
Post FLIR				....

*** $p < .001$

For the Fixed-KR condition, correlations among the first-sesssion training scores and sessions to pass vehicle sets were computed. Most correlations were not significant. Only two significant. Both were negative, indicating that the higher the initial score, the fewer sessions required to pass (see Table C-14).

Correlations between the number of sessions required to reach the 85% criterion and other scores differed for the two training conditions (see Table 2-12). For the Self-KR+ group, number of sessions correlated significantly with pre- and posttest scores and with the first-session score of the all-vehicle set. This was not the case for Fixed-KR. These different patterns probably reflect the effect of the prior part-task training for the Fixed-KR group. Their initial skill was less important in determining first-session scores than was the case for Self-KR+.

Table 2-12
Experiment 2: Correlations Between Number of Sessions to Training Criterion and Pre-Posttest, Training and Transfer Scores - Pass85 Sample

	# Sessions to Criterion on All-Vehicle Set		
Score	All soldiers	Fixed-KR	Self-KR+
Pre- and Posttests			
Pre Visible	-.09	.25	-.72*
Pre FLIR	-.14	.18	-.56
Post Visible	-.09	.14	-.62
Post FLIR	-.05	.18	-.74*
Training Score on First Session of All-Vehicle Set	-.84	-.32	-.97**
Transfer Test Scores			
No Transfer	-.24	-.12	-.37
Day Black-Hot	-.10	-.12	-.25
Night Black-Hot	-.20	-.30	-.38

*$p < .05$ **$p < .000$

Table 2-13 presents correlations between the pre- and posttest scores, transfer scores, and initial score on the all-vehicle training set. Pre- and posttest scores correlated similarly with the other scores. The same correlation pattern was expected, given the high correlations among the pre- and posttests. Of particular interest was the different relationship with the first-session score on the all-vehicle training set as a function of training condition. For the Fixed-KR group, pre- and posttest scores did not correlate with the first-session score. But for the Self-KR+ group, pre- and posttest scores did correlate with the first-session score. Again, the lack of a relationship for the Fixed-KR group probably reflects the effect of the part-task training which preceded the all-vehicle set for this group of soldiers

The pre- and posttest scores also correlated significantly with the subscores on the transfer test. The correlations among the transfer subscores themselves were also significant. For the Fixed-KR group, training scores on the all-vehicle set did not correlate with transfer

scores, while training scores for the Self-KR+ group did correlate with transfer. Overall, the results show that the most consistent predictor of individual scores on transfer was initial skill.

Table 2-13
Experiment 2: Correlations Between Pre- and Posttest, Transfer, and Training Scores for the Entire Sample

Score	Training Score			Transfer Score		
	First Session All-Vehicle Set:			No Transfer	Day Black-Hot	Night Black-Hot
	All Ss	Fixed-KR	Self-KR+			
Pre- and Posttests						
Pre Visible	.38*	-.04	.77***	.58***	.66***	.60***
Pre FLIR	.45**	-.01	.70**	.67***	.69***	.64***
Post Visible	.43*	.12	.72**	.61***	.63***	.57***
Post FLIR	.43*	-.07	.81***	.64***	.71***	.75***
Training Score on First Session of All-Vehicle Set		...				
All Ss				.72***	.33	.52**
Fixed-KR				.37	.13	.15
Self-KR+				.76***	.62**	.72**
Transfer Score						
No Transfer					.73***	.79***
Day Black-Hot					...	.71***
Night Black-Hot						...

$^*p < .05$ $\quad ^{**}p < .01$ $\quad ^{***}p < .001$

Experiment 2: Discussion

Answers to the Questions

- Is a training strategy that requires multiple, timed vehicle identification trials and provides knowledge of results more or less effective than a training strategy, with no time requirement that also provides corrective visual feedback?

In general, the findings support the conclusion that the self-paced training with corrective visual feedback was the better training strategy. The part-task training for the soldiers in the Fixed-KR condition clearly benefited them when they began the all-vehicle training set. Their identification scores were significantly higher on the first session compared to those in the Self-KR+ condition. However, there were no differences between the two groups in the number of sessions required to reach criterion on the all-vehicle set. A detailed analysis of soldiers who experienced difficulty in reaching criterion on the all-vehicle set showed that those in the Self-

KR+ group tended to have a more steady improvement in performance with training than those in the Fixed-KR group. This occurred despite these Self-KR+ soldiers having lower initial scores on the all-vehicle set and lower pretest FLIR scores.

For the transfer conditions, those with Self-KR+ training scored higher by about 9 percentage points than those with Fixed-KR training. No differences between the two training strategies occurred for the no-transfer conditions. On the posttest, for the Pass85 sample, soldiers in the Self-KR+ group improved significantly more from pretest to posttest FLIR imagery scores than those in the Fixed-KR group (13% points difference). Finally, the need to provide a means of exiting a training exercise when a learning plateau is reached was demonstrated with both training strategies.

- Is transfer greater to the same vehicles displayed in black-hot day imagery or black-hot night imagery, regardless of training strategy?

For both training conditions, vehicle identification scores were significantly higher, by 9 percentage points, for night, black-hot imagery than for day, black-hot imagery.

- Does practice in determining aspect angle improve aspect scores in a transfer condition?

Practice in determining aspect resulted in significantly higher scores on the transfer test as compared to soldiers who did not have the benefit of such training. Aspect scores were 13 percent higher for those in the Self-KR+ group, who had practice in determining aspect angle in the Vehicle ID module, compared to those in the Fixed-KR group, who were not required to determine aspect angle during training.

Secondary questions were:

- Does thermal imagery training transfer to visible imagery, regardless of training strategy? Does thermal imagery training transfer to vehicles not trained, regardless of training strategy?

Scores for vehicles in the FLIR training increased from pretest to posttest (20 percentage points); scores did not change for vehicles that were not trained. This was the case for both visible and FLIR imagery. Time to respond to the FLIR images also improved: twice as fast on the posttest as on the pretest. In contrast, time to respond to the visible images did not change.

- What vehicles tend to be confused with each other during training and testing?

During training, initial confusions were primarily among the tanks and among the APCs. The T72 and the M1 were more likely to be confused with each other than with the M60. The BTR and the LAV were confused with each other. During transfer these confusions persisted, but were not as frequent. As would be expected, fewer confusions occurred on the posttest than on the pretest. Many confusions were eliminated. For example, the M60 was no longer

identified as the LAV or the M551. The T72 was no longer misidentified as the LAV, BTR, M113, or ZSU. Of the vehicles trained, however, some confusions remained. The LAV was misidentified as the M2 and the BTR; the BTR misidentified as the LAV. For the vehicles not trained, the BMP and ZSU tended to be confused with a variety of other vehicles. Clearly, without training soldiers did not know these vehicles.

The Starting Point

Two groups of soldiers participated in Experiment 2. Students from the BLC were lieutenants with few years of Army experience. The NG participants were noncommissioned officers, privates and specialists. Over half had been in the Army for at least 11 years. In addition, the two groups did not begin the training with the same level of vehicle identification skill. The BLC students participated in the experiment when they were in the middle of combat vehicle instruction. This contributed to higher pretest scores (about 35% points higher) than the NG participants. Despite this difference and because of the small sample size within each group, the two groups were pooled for the data analyses.

The Learning Process

The primary measures of interest during the practical exercises were the number of sessions required to meet the training criterion on the all-vehicle set, vehicle identification scores, vehicle confusions, total training time, and response times to each image display. These measures were used to determine which training strategy was better and to identify needed improvements to the CVI*Plus* program.

As time constraints and computer problems prevented all soldiers from reaching the original criterion of 91% during the all-vehicle training set, the criterion was changed to 85% for this set. Not all soldiers reached this criterion, however. Those who achieved 85% were called the Pass85 sample and provided a similar basis for group comparisons, as all in this sample had achieved the same minimum level of proficiency.

As in Experiment 1, the results and our observations of the training indicated that a program with remedial training capability, and tailored, diagnostic feedback at the end of a training session would probably have made learning more efficient, particularly for some in the Fixed-KR group. Although not significant, proportionately more of the Self-KR+ group did not achieve the 85% criterion on the all-vehicle training set. Both the non-timed feature of the Vehicle ID module and the inability to create a sequence of part-task vehicle sets with this module contributed to the longer training times with the Self-KR+ group.

Two other factors impacted the learning rate. One factor was distance to the vehicle. All the training and transfer images were presented at the next to the farthest range in the program (Range 2). Distance to vehicles, as shown in prior research (e.g., with scaled models in Haverland & Maxey, 1978) affects learning progress. O'Kane et al. (1997) found lower scores with thermal images at simulated ranges of 1 and 2 km. If the next closest range had been used, it is likely that more soldiers would have reached criterion and reached it sooner. The other

factor was the high criterion. The 91% criterion meant that a soldier could make no more than 4 errors out of 48 images. Meeting this criterion required a high level of concentration. In Signature Challenge, which is timed, it is easy to make an error accidentally, and there is no opportunity to correct such errors. Responding quickly and making accidental errors was typical of the soldier in the Fixed-KR condition who required 19 sessions to reach criterion.

Soldiers in the Fixed-KR condition varied considerably in the number of sessions required to meet criterion on each of the vehicle sets. On the part-task vehicle sets, the number of sessions ranged from 1 to 11 and 12. Pretest FLIR scores correlated significantly with the number of sessions on the tank set, $r = -.86$, $p < .0000$, but not with number of sessions on the APC set, $r = -.31$, $p < .22$. The point was made previously that remedial training and diagnostic summary feedback would have helped soldiers on the all-vehicle set. These features would probably help reduce the number of sessions on the smaller vehicle sets as well.

With regard to vehicle scores, it was not surprising that soldiers in the Fixed-KR condition scored significantly higher on the initial session of the all-vehicle set than those in the Self-KR+ session (82% versus 68%). On average, those in the Fixed-KR condition had been trained on 48 images of each tank and on 32 images of each APC prior to the all-vehicle set. Those in the Self-KR+ condition had no practice with these images prior to the all-vehicle set. Significantly more in the Fixed-KR condition reached the 91% criterion, but not the 85% criterion.

For Fixed-KR, the part-task training reduced initial vehicle confusions on the all-vehicle set. On the first session of this vehicle set, identification scores on each vehicle for the Fixed-KR group were higher than the corresponding scores for the Self-KR+ group. In addition, cross-category tank and APC confusions were less for the Fixed-KR group, which again reflects the impact of their prior part-task training.

Response times to each display provided additional insights into how the formats of the Signature Challenge and Vehicle ID modules influenced soldier behavior. As expected on the all-vehicle set, soldiers in the Self-KR+ condition spent more time on each trial than those in the Fixed-KR condition. Self-KR+ times were ten times longer than Fixed-KR times (28.33 sec vs. 2.65 sec). The Self-KR+ condition required two responses, vehicle and aspect, and there was no time requirement. The Fixed-KR condition required only one response, and the maximum exposure time was seven sec for the part-task sets and ten sec for the all-vehicle set. However, the requirement to make two responses in Self-KR+ cannot account for the ten-fold increase in time per trial. Soldiers in the Self-KR+ condition took advantage of the opportunity to study each FLIR image. Some would determine aspect first and use aspect to help identify the vehicle. Additional time was also spent in examining the corrective visual feedback when errors were made.

The total average times spent on the SC and Vehicle ID modules were quite similar, 54 and 59 minutes respectively. However, there were considerable individual differences in training time. In each condition, the slowest individual required 5.5 times longer than the fastest

individual. For Self-KR+, this time difference was 97 minutes. For Fixed-KR, the difference was 75 minutes.

In general, it would appear from the findings and from observations of soldiers during training, that some tended to respond too fast in Fixed-KR. On the other hand, some tended to take too much time in Self-KR+. Both approaches have negative consequences during initial discrimination learning. When responding very quickly in Fixed-KR, the soldier does not take advantage of the time allowed to study the images, resulting in errors. But when soldiers study a vehicle for a very long time (e.g., 30 to 60 sec) in Self-KR+ because they are uncertain, little is gained by this scrutiny of the vehicle. Time would be better spent by responding with their best "guess." When they err, they should then take time to examine the corrective visual feedback to help identify vehicles.

A trend analysis of the ten soldiers who required more than three sessions to reach the 85% criterion on the all-vehicle set indicated that the Self-KR+ condition provided the better learning environment for these soldiers. Those in the Fixed-KR condition tended to require more sessions to reach criterion, yet they had the higher pretest FLIR and first-session scores on the all-vehicle set. Learning curves also showed a more steady progression toward criterion for those in the Self-KR+ condition, than those in Fixed-KR.

Findings from both training conditions pointed to the need for remedial training when soldiers do not reach criterion within approximately three sessions. On-site observations also showed the need to provide summary information on vehicle confusions at the end of each session. As in Experiment 1, soldiers did not necessarily remember or know which vehicles they confused. Such feedback would pinpoint these confusions, and if remedial training approaches were incorporated as well, learning should be more efficient and provide soldiers a better understanding of the vehicles' thermal signatures.

In conclusion, the answer to whether the Fixed-KR approach, which provided many, short exposures to images, was better than a Self-KR+ approach, which provided for self-study and immediate visual corrective feedback but took more time, was not clearly answered from the training results. The transfer data and the pre-posttest results provided additional insights on this issue.

The Outcome

The diversity of thermal images in a training program is limited. But FLIR imagery in the real world is limitless, and it is very dynamic. The imagery depends on the temperature, the operating time of the vehicle, the surrounding terrain, atmospheric conditions, the contrast and brightness settings on the system's sight, and other factors. So the ability of soldiers to transfer skills could be viewed as even more important than learning a specific set of images. Clearly, there was a need to examine the degree to which instructional strategies enhanced transfer, even if it was limited to new imagery within the program itself.

Findings from the transfer test helped clarify the effects of the two training strategies and provided important data on the ability of soldiers to interpret new imagery. To reiterate, the transfer test had three components: no transfer images (white-hot night) and transfer with black-hot night images and white-hot day images. In addition, the transfer test assessed ability of soldiers to determine the vehicle's aspect angle.

Scores on the no transfer images showed no significant differences between the two training conditions for either the entire sample or the Pass85 sample. Yet, for the entire sample, the Fixed-KR score was 15 percentage points higher than the Self-KR+ score. This probably reflects the fact that there was a trend for more of all the Fixed-KR participants to reach criterion during training than was the case for the Self-KR+ participants. For the Pass85 sample, there was only 2 percentage points difference between the two groups. The smaller difference was accounted for by the increase in transfer scores for the Self-KR+ group.

As expected, transfer was greater to black-hot night images than to black-hot day images. The black-hot night scores were significantly higher than black-hot day by 9 percentage points, for both the entire sample and the Pass 85 sample. Also, the Pass85 sample scores were higher than scores for the entire sample, by 8 percentage points, for both black-hot conditions.

As discussed previously, one explanation of the differences between the black-hot night and day scores may be explained by the theory of identical elements and Osgood's (1949) transfer surface. Cormier (1987) summarized more recent research on encoding and retrieval processes that provide another, yet similar, way of conceptualizing these effects. Of relevance to vehicle discrimination is the concept that transfer is a function of the way the original material is encoded and of the cues or information available at the time of transfer. To the degree that information encoded during learning is retrievable with the cues available in the transfer situation and the transfer situation contains critical cues that help reinstate the original stimulus situation, then positive transfer will occur. But if the encoding of the original material is incompatible in some way with information or material in the transfer task, then less transfer will occur. From this perspective, the thermal cues presented during training of the white-hot night images should have been retrievable from the black-hot night images, as the instruction on what parts of each vehicle were hot applied regardless of whether the imagery was white-hot or black-hot. The exercises provided additional experience in discriminating vehicles in accordance with these thermal cues. On the other hand, no training was provided on the thermal cues that appear during the day. Even though some cues apply to both day and night, daytime thermal cues can be quite distinct. Thus what was learned on the original task of white-hot night imagery was, in some instances, inconsistent with the black-hot day imagery shown on the transfer test.

No specific hypotheses were made regarding how the training conditions would affect transfer. Yet for the Pass85 sample, the Self-KR+ group scored significantly higher than the Fixed-KR group, 88% versus 79%, on the black-hot imagery. This occurred despite the fact that the Fixed-KR group had 2.5 more trials during training than the Self-KR+ group, 312 trials compared to 125, respectively. Thus the repeated and short exposures to vehicle images in the Fixed-KR condition did not result in higher transfer scores. This finding raises the question of what soldiers were learning in the two training modules. Were soldiers in Fixed-KR treating the

training like a test, attempting to "memorize" many of the images, and therefore performing relatively elementary processing operations? Were soldiers in Self-KR+ using the self-study time available and the corrective feedback to gain an understanding of the vehicles' thermal signatures and what key features discriminated the vehicles they tended to confuse with each other? Did the requirement for soldiers to determine aspect angle in Vehicle ID also cause them to process the imagery at a deeper level? Many soldiers determined the aspect first and then used that information to help them identify the vehicle.

The point has been made in the literature on transfer (Cormier, 1987; Cormier & Hagman, 1987; Lupart, 1995; Marini & Genereux, 1995) that there is frequently a trade-off between training for transfer and training for rapid acquisition of the training tasks. If the training tasks are the target, then training is more rapid, as fewer task elements are typically involved. If transfer is the target, then training not only involves the training task but preparing individuals for transfer as well (Cormier & Hagman, 1987). Salomon and Perkins (1989) differentiated between the "low road" to transfer based on extensive practice and automatization on training tasks and the "high road" to transfer based on mindful abstraction and deep understanding of training tasks. They claim that although low-road learning can result in fast attainment of transfer, the resulting transfer can be limited in scope and flexibility. In contrast, high-road learning results in slower attainment of transfer, but the transfer attained is more extensive and adaptive in nature. Salomon and Globerson (1987) argue that proximal factors, such as the nature of the training materials and instructional procedures, can influence the learner's approach to a task and whether a mindful approach is taken They cite research on how instructional procedures (e.g., view a TV show "to learn" or "to have fun," learn with the intent to explain the material to others, generate images of the material and relate it to what you already know, probe students about what they are learning) can influence what is learned and how individuals approach tasks.

Although we cannot claim automatization of skills with Fixed-KR, nor a deep understanding of thermal signatures with Self-KR+, there are parallels which may help explain the difference in transfer scores. Soldiers in Fixed-KR had more exposures to the same training images than soldiers in Self-KR+, and they had to respond quickly. The format of Signature Challenge may have signaled to those in the Fixed-KR condition that the important factor was to identify vehicles quickly. The time limit may have forced automatic processing rather than controlled processing of the images. In Self-KR+, soldiers may have executed a more controlled processing of images as the ten-fold increase in time to respond would indicate. For some soldiers, this increased time could indicate self-regulation of their learning and use of the feedback to enhance understanding of thermal signatures of vehicles; techniques that have been shown to enhance transfer (Lupart, 1995). Conversations with many soldiers, as they worked in the Vehicle ID module, indicated they did use the time to study the image, determine its aspect, and relate what they saw to the thermal cues presented by the instructor. For these soldiers, this was a conscious process, a deliberate mental effort while learning the vehicles, consistent with Salomon and Perkins' (1989) definition of mindfulness.

Lastly, the transfer test scores for vehicle aspect differed across training conditions, with flanks being the easiest, followed by rear obliques, and then front oblique. This occurred for the

entire sample and the Pass85 sample. Given that determining aspect is an easier task than naming vehicles, we had not necessarily expected training on aspect to have an effect. But the findings did show positive effects from such training, with the Self-KR+ scores on aspect 12 percentage points higher than Fixed-KR scores.

The pretest-posttest results paralleled those in Experiment 1. The scores for vehicles in the FLIR training increased about 20% points from pretest to posttest for the entire sample and for the Pass85 sample. Scores for vehicles that were not trained remained the same. This pattern occurred for both visible and FLIR imagery. Enhanced skill and confidence with the FLIR imagery were also demonstrated by the reaction times to each display on the posttest. Times were twice as fast on the posttest, and approached reaction times to the visible imagery.

In this experiment, pre-posttest changes were affected to some degree by the training conditions. First, further evidence for the effectiveness of the Self-KR+ strategy was shown on the posttest. For the Pass85 sample, posttest scores on all FLIR images (both trained and not trained vehicles) increased more for the Self-KR+ group than the Fixed-KR group, 18% and 5%, respectively. Second, the change in reaction times on the posttest may also reflect the response rates and styles developed during training. For the vehicles that were not trained, Fixed-KR times decreased more than was the case for Self-KR+ times.

Training to criterion had positive effects. The Pass85 sample was more proficient than the entire sample. On all transfer scores and on the posttest scores, the Pass85 sample tended to score higher than the entire sample. In addition, the Pass85 sample responded faster on the transfer test and posttests.

In general, scores correlated with each other; individual differences in skill were maintained throughout the course of training and testing despite an increase in proficiency for the sample as a whole. Pretest and posttest correlations were high, and they correlated with transfer scores. The number of sessions to achieve criterion on the all-vehicle training set correlated with the initial score on that set. Yet, the pretest and posttest scores related differently to performance measures on the all-vehicle set for the two training conditions. Significant relationships with the pretest occurred for the Self-KR+ group only. This finding probably reflects the impact of skills acquired by the Fixed-KR group during the part-task training, and the absence of part-task training, prior to the all-vehicle set, for the Self-KR+ group. Consequently, initial (pre-training) skills for the Fixed-KR group were less important in predicting the rate of progress and scores on the final all-vehicle training set than was the case for the Self-KR+ group.

In conclusion, if part-task training could have been incorporated in the Self-KR+ condition, a more direct test of the effects of each training strategy would have resulted. Nonetheless, the transfer and pre-posttest findings indicated that the self-paced training with visual corrective feedback was a better training strategy than fixed-paced training with knowledge of results, for initial learning as well as transfer. We want soldiers to identify vehicles quickly on the battlefield. The findings from Experiment 2 indicate that the appropriate learning sequence may be first giving soldiers the skill to interpret the thermal signature of vehicles, and then training them to respond quickly.

The results from this experiment do not imply that Self-KR+ is the only training strategy that can enhance transfer, nor that Fixed-KR is an undesirable training procedure. In Experiment 3, we combined the Self-KR+ and Fixed-KR training in order to take advantage of the training features in both the Vehicle ID and Signature Challenge modules.

Experiment 3: Training at Near and Far Ranges

Experiment 3: Background

The third experiment was designed to determine the effects of training soldiers to identify vehicles at near versus far ranges. Also included was an examination of how these skills transferred to vehicles presented at "non-trained" ranges, specifically whether skill in identifying vehicles at near ranges transferred to skill in identifying vehicles at far ranges, and vice versa. The point has been made in previous research on vehicle recognition that training on cues seen primarily at close distances may be counterproductive in learning the critical cues visible at tactical ranges or under degraded conditions (Cockrell, 1979; Warnick & Kubala, 1979; Warnick & Smith, 1989). In fact, much of the photopic imagery available for vehicle and aircraft recognition programs is at close range (e.g., DA, 1984a, 1987; Foss, 1995; Foss & Gander, 1996; Jane's Information Group, 1997). The Army's armored vehicle recognition training aid, GTA 17-2-13 (DA, 1984b), is a series of flashcards with line drawings of vehicles. Although small in scale, the detail in these drawings is much greater than that visible at tactical distances. Warnick and Kubala (1979) attributed poor pretest performance in identifying vehicles at a distance to soldiers being trained to look for cues not visible at long ranges or to look for cues that were not always present. They concluded that the soldiers' previous training with close-up imagery had provided too much detail.

On the other hand, there is an argument for training at close distances so soldiers thoroughly understand the heat plant and the structural characteristics that generate the vehicle's thermal signature. Such understanding may also provide the conceptual skills necessary for interpreting thermal signatures of vehicles at farther ranges and under different atmospheric conditions. Differences in the thermal signature of vehicle components are more easily described and explained with close-up than with far views. Wallis (as cited in Smith et al., 1980) stated that individuals should have full and complete exposures of a vehicle, including flank and front views, in order to build a mental picture of it.

In Experiment 3, the near range band included the Close-up Range and Range 1 in CVI*Plus*. The far range band included Ranges 2 and 3. The training sequence reflected what was learned about training CVI skills from Experiments 1 and 2. Six vehicles were trained. Group instruction on thermal cues for each vehicle was first. Initial practice on identifying vehicles was conducted with the Vehicle ID training module. Signature Challenge training followed. Vehicle ID preceded Signature Challenge in order to reduce the learning plateaus that occurred previously with Signature Challenge. Soldiers were instructed to take their time in Vehicle ID and to study the corrective visual feedback. The intent was to have them understand why their responses were incorrect, thereby reducing vehicle confusions.

The number of images displayed in both Vehicle ID and Signature Challenge was reduced. The long vehicle sets in the previous experiments were mentally fatiguing for some soldiers, requiring long periods of concentration in order to avoid errors. The longer sets frequently lowered their motivation as well. In Signature Challenge, once some soldiers found

they had made more than the allowable number of errors, they simply selected any response to get through the imagery set and begin again. With the single large Vehicle ID set in Experiment 2, which was not timed, some soldiers took a long time to progress through the training. The transfer tests were also shortened to maintain motivation during the testing process.

Vehicle ID training was limited to one set of vehicle imagery, as Vehicle ID was not gated in the CVI*Plus* version available. To shorten the training exercise, the four cardinal views, not all eight aspects, of each vehicle were presented.

The vehicle sets in Signature Challenge training were limited to two and three vehicles. The Signature Challenge sequence culminated in two sets that presented new groupings of the six vehicles. The previous experiments had shown the value of such training. For example, training a soldier to discriminate the M1 from the T72 with a high degree of accuracy did not guarantee that he could then automatically discriminate the M1 from the ZSU with the same degree of accuracy. The vehicle sets in the Signature Challenge training exercises required discrimination among the vehicles most often confused.

Both Vehicle ID and Signature Challenge tests examined transfer of skill to another range. The previous experiments had relied on the Vehicle ID test only. It was deemed essential to determine whether transfer results would be similar, regardless of the test format.

Experiment 3: Purpose

The primary questions addressed in Experiment 3 were:

- Is the amount of training required to identify vehicles at near ranges the same or different from the amount of training required to identify vehicles at farther ranges?

- Does skill at identifying vehicles at near ranges transfer to farther ranges?

- Does skill at identifying vehicles at far ranges transfer to nearer ranges?

The secondary questions were:

- Does thermal imagery training transfer to visible imagery regardless of the training condition?

- Does thermal imagery training transfer to vehicles not trained regardless of the training condition?

In investigations of how far away aircraft or vehicle features can be discriminated in daylight, researchers (e.g., Gibson, 1947; Foskett, Baldwin, & Kubala, 1978; Heuckeroth, Smith & Warnick, 1988b) simulated how aircraft or vehicles would appear at actual distances. Gibson (1947) presented a good discussion of the rationale for this technique, and its advantages and

disadvantages. In these research efforts, scaled models, photographs, or projected images were presented. Observers in these studies typically used either their eyes or a tactical sight. The distance of the observer to the scaled models or images varied with the scale of the model as well as the tactical distance of interest.

One critical factor is that simulations must replicate the visual angle subtended by the object of interest at the tactical distance of interest. For aircraft, Gibson (1947) stated that

> it was at least possible to assume that the identifiability of a model seen against a background of sky at a considerable distance was approximately the same as the identifiability of a real plane seen against the same sky at a proportionately greater distance. The model and the real plane would intercept the same visual angle on the retina of the eye; both would be approximately silhouettes; and the details of both images would be equally just barely noticeable (p. 160).

Gibson used 1 to 72 scaled plastic models of aircraft, with observers standing at distances that stimulated tactical distances ranging from half a mile to four miles. He found that the ability to recognize an aircraft depended on its distance from the observer and its attitude; that the head-on position was usually the most difficult; and that identifiability depended on aircraft size. As discussed by Gibson, a problem with this procedure is that atmospheric haze and the resulting diffraction of light waves will reduce the identifiability of real plane compared to a model. So observers probably identified aircraft at much farther ranges than they would in an operational environment.

We expected similar effects with vehicles and with FLIR imagery. The near imagery in CVI*Plus* program was expected to be easier to learn than the far imagery, as more distinguishing cues are visible at the near ranges. Prior research supported this expectation. Heuckeroth et al. (1988b) used thermal photographs of vehicles, obtained with 1st generation FLIR sights at three range bands (800-1200 m, 1300-1700 m, and 1800-2300 m). Imagery was of oblique and flank views only. These images were reproduced on 35mm slides for training and testing. M1 tank and M2 Bradley crewmen were trained to a criterion of 80% at the closest range band, and then tested on their ability to identify the same vehicles at the three range bands. Test scores were 79% at the closest range band, and dropped to 48% at the two farther range bands. The authors found, however, that some vehicles had characteristics that remained discernable at longer ranges, and that shape was the primary distinguishing feature used in identification. Unfortunately, the power setting on the thermal sights used in taking the photographs was not specified, nor how the visual angle subtended by the image at the tactical ranges was replicated on the 35mm slides.

In an earlier effort with models of armored vehicles, Foskett et al. (1978) had individuals name features of the vehicles they saw with their unaided eye at simulated ranges from 100 m to 4000 m. Although the features cited depended in part on the individual's expertise, the size of the feature was also important. At distances beyond 1200 m the primary features discerned were tracks versus wheels, presence of a turret, and turret location. Smaller features such as multiple gun tubes, sprocket locations, and number of road wheels were seen at ranges less than 800 m.

In CVI*Plus*, vehicles were presented at four ranges: Close-up, Range 1, Range 2, and Range 3. These ranges were achieved by photographing the vehicles at different ranges and with different fields of view through the thermal camera. If viewed through a thermal sight with unity power, these ranges, as presented on the computer screen, translated into approximately 65-75 m, 130-140 m, 195-205 m, and 260 to 270 m, respectively. Thus the images at the Close-up Range were four times the size of those at Range 4, three times larger than images at Range 3, and twice the size of those at Range 2. In contrast to prior research with scaled models where no degradation of image resulted from atmospheric conditions, the images did depict the appropriate spatial degradation for each range. In other words, the image at Range 4 was not simply a reduction of the Close-up Range image. Instead, because the images presented were "real" imagery, the effects of range upon the image as seen by a soldier using a thermal sight were displayed in the training program.

No specific hypothesis was made regarding how much more training would be required to identify vehicles at the far range band. Haverland and Maxey (1978) did find, however, that helicopter crewmen required about 1.5 times as many presentations to identify models of vehicles to a criterion of 90% at the simulated range of 4000 m than at 3000 m when looking through a 13x day sight. These distances correspond to 308 m and 230 m, respectively, for the unaided eye. However, degradation resulting from atmospheric haze was not simulated or portrayed in their research.

Soldiers who trained on the near imagery were expected to have some difficulty transferring their skills to the farther range. However, those who trained on the far imagery were expected to do well at the near range. It was also assumed that soldiers would maintain their proficiency on the no transfer imagery in these tests. Haverland and Maxey (1978) found no differences in vehicle identification test scores at the simulated distances of 3000 or 4000 m. Scores were the same whether soldiers trained at 3000 m or 4000 m. However, a powerful 13x tactical sight was used in this experiment. When these ranges are converted to the distances as seen with the naked eye, the vehicles would be about 75 m apart. Given this minimal separation, together with no simulation of atmospheric degradation, it is not surprising that Haverland and Maxey did not find differences as a function of range.

In Experiment 3, individual practice in discriminating vehicles at the far and near range bands was assumed to have a stronger effect than group instruction at these two range bands. In the group instruction phase, soldiers are passive. They listen to an instructor describe key features of each vehicle as the images are presented on a large screen. In the practice or training exercise phase, soldiers are active. All soldiers were to have a minimum of 168 computer-based practice trials; they could progress at their own pace; feedback tailored to their individual responses was given on each trial.

No strong expectations were held regarding whether the two test formats in CVI*Plus* would yield differences in scores. The primary distinction between the two was that Signature Challenge was timed and allowed no change of response, whereas Vehicle ID was not timed and

soldiers could change their response before the next image appeared. Signature Challenge, therefore, did place more pressure on the soldier.

Based on the previous experiments, it was expected that FLIR training would have positive effects, as indicated by pretest-posttest differences, on both the FLIR and visible images of the vehicles in the training. These skills were not expected to transfer to the other vehicles.

Experiment 3: Method

Participants

The soldiers were from a Mechanized Infantry (Bradley) Battalion. Of the 67 who participated, 35% were Sergeants or above, 24% were either Specialists or Corporals, 37% were Privates, and one was a Second Lieutenant. They were between 18 and 33 years of age, with an average age of 23.12 ($SD = 3.24$). One-fourth had a year or less of military experience, 34% had between two and three years, 28% between four and six years, and 12% had seven or more years of experience.

Over half (51%) were either Bradley commanders or gunners, 28% were drivers, and 12% were from the dismount squad. One platoon leader, one platoon sergeant, and three soldiers from the battalion S-3 also participated. Soldiers varied in the time spent in their current duty positions. One-third were new to their position, with less than one month experience, and another one-third had between six and twelve months in their current position. Twelve percent had from 13 to 18 months experience, and 21% had 19 months or more experience.

In general, most soldiers had experience with thermal sights, but had little formal or specialized training in combat vehicle identification with thermal sights. Some (16%) had never used a thermal sight; 30% had used more than one thermal sight. The majority (79%) had used the Integrated Sight Unit on the Bradley; 30% had used the Dragon Night Tracker; 12% had experience with the TOW thermal sight. Half the soldiers had never used a thermal sight in a tactical situation. The remaining soldiers had employed thermal sights in Desert Storm, Jordan, Bosnia, and/or at a Combat Training Center. Twelve percent had received special vehicle identification training in a field exercise five months prior to the experiment.

Design

Near and far FLIR imagery was varied during both instruction and training. Near images were defined as those at the Close-up Range and Range 1; far images, Ranges 2 and 3. Soldiers first received group instruction on either near or far imagery. Then soldiers in each instructional group were divided in half. Half practiced with near images in Vehicle ID and Signature Challenge; the other half practiced on far images. Soldiers were tested for transfer with imagery presented at the ranges not used in the practice sessions. Pretests and posttests were also given. Four hours were allowed for each experimental session. The experiment was executed in five

consecutive days with two iterations of the design per day. An outline of the design and sequence of events is in Table 3-1.

Table 3-1
Design of Experiment 3 and Sequence of Events

Sequence of Events	Description			
1. Pretests	FLIR test, then Visible test.			
2. Group Instruction on Vehicles and Vehicle Cues	Overhead projection using Image Library. - Vehicles were: M1, T72, M2, BMP, LAV, BTR. - Each vehicle presented on the 2-display screen with comparisons of white-hot night with black-hot night images. Each aspect displayed. - Soldiers randomly assigned to Near Instruction (vehicles at Close-up Range) and to Far Instruction (vehicles at Range 2).			
	Near Instruction ($n = 35$)		**Far Instruction** ($n = 32$)	
3. Test on engine and exhaust locations	Paper-pencil test.			
4. FLIR Practice	Random assignment of soldiers within the Near and Far Instruction groups to training exercises using either Near or Far FLIR imagery. - Near Practice with Close-up and Range 1 images. - Far Practice with Range 2 and Range 3 images. - All images were white-hot, night.			
4a. Vehicle ID Exercises (Table 3-2)	**Near Practice** ($n = 17$)	**Far Practice** ($n = 18$)	**Near Practice** ($n = 16$)	**Far Practice** ($n = 16$)
4b. Signature Challenge Exercises (Table 3-3)	**Near Practice** ($n = 17$)	**Far Practice** ($n = 18$)	**Near Practice** ($n = 16$)	**Far Practice** ($n = 16$)
5. Transfer Tests (Table 3-4)	36 FLIR displays in each test. All images were white-hot, night. - Oblique views tested transfer to ranges not in the training exercises. - Front and rear views were no transfer images.			
5a. Vehicle ID Test	Transfer to Far	Transfer to Near	Transfer to Far	Transfer to Near
5b. Signature Challenge Test	Transfer to Far	Transfer to Near	Transfer to Far	Transfer to Near
6. Posttests	Same as Pretests.			

Note. Experimental conditions are in bold font. Images were presented randomly within each exercise and test.

Group instruction was given to seven soldiers simultaneously; the number corresponding to the number of computers available. After this instruction, soldiers were randomly assigned to practice on either near or far imagery. Therefore for soldiers in two of the conditions, instruction and practice were consistent. All images were at either the near or the far range band. For

soldiers in the other two conditions, instruction was at one range band and practice at the other. The number of soldiers within each instruction-practice combination ranged from 16 to 18.

Soldiers trained on six vehicles: two tanks (M1 and T72), two tracked APCs (M2 and BMP), and two wheeled APCs (LAV and BTR). Three vehicles were US; three were nonUS.

Group instruction. The group instruction was conducted by a military instructor; the same instructor as in Experiments 1 and 2. An overhead projection system was used to display the Image Library module. Each vehicle was presented with the two-display screen. One display had white-hot night imagery; the other had black-hot night imagery. The specific range presented was Close-up for the Near instruction; Range 2 for the Far instruction. Thus for each instruction condition, the more detailed and the closer, imagery was used. Vehicles were presented in the following sequence: M1, T72, M2, BMP, LAV, and BTR. Each aspect of each vehicle was presented in the following order: right front oblique, front, left front oblique, right flank, left flank, left rear oblique, rear, and right rear oblique. This instruction took approximately 30 minutes.

After group instruction, soldiers were given a multiple-choice test on the engine and exhaust locations on each vehicle (see Appendix D). From the prior experiments, it appeared that soldiers who had not mastered this material initially had difficulty in identifying vehicles and progressed more slowly during training. The test examined this issue more closely.

Vehicle ID exercises. Vehicle ID training was the first training exercise. It was not timed, allowing soldiers to study the corrective visual feedback provided when they erred. This training was included to reduce the plateau effects that occurred on the Signature Challenge training exercises for some soldiers in Experiments 1 and 2. The number of displays was 24, with only the cardinal views of each vehicle presented. Within each practice condition, half the displays were at the closer of the two ranges within each range band; half at the farther range. Details on these exercises are in Table 3-2. Soldiers were told to learn as much as they could from the corrective visual feedback before they progressed to Signature Challenge. Soldiers went through Vehicle ID once; no pass criterion was imposed. Feedback on the correctness of the vehicle and aspect responses was given after each trial. Summary feedback consisted of percent correct vehicle identifications and percent correct aspect responses.

Table 3-2
Experiment 3: Format of Vehicle ID Training Exercises

	Practice Condition	
Vehicle Aspect Angle	Near Range Band	Far Range Band
Front	Close-Up	Range 2
Rear	Range 1	Range 3
Left Flank	Close-Up	Range 2
Right Flank	Range 1	Range 3

Signature Challenge exercises. The Signature Challenge training sequences are summarized in Table 3-3. For each practice condition, the first three vehicle sets paired vehicles at the closer range in the range band. The two tanks (US and nonUS) were paired, the two tracked APCs were paired, and finally the two wheeled APCs were paired. The next three sets (4 through 6) repeated this sequence, but at the farther range in each band. The last two sets grouped the US vehicles and the nonUS vehicles; all were presented at the farther range. The training was progressive in that the closer range, and presumably the easier, was presented first. The training sequence also allowed more practice at the farther, more difficult range within each range band.

Table 3-3
Experiment 3: Format of Signature Challenge Training Exercises

Set #	Vehicles Presented	Range Band	
		Near Practice	Far Practice
1	M1 & T72	Close-up	Range 2
2	M2 & BMP	Close-up	Range 2
3	LAV & BTR	Close-up	Range 2
4	M1 & T72	Range 1	Range 3
5	M2 & BMP	Range 1	Range 3
6	LAV & BTR	Range 1	Range 3
7	M1, M2 & LAV (US)	Range 1	Range 3
8	T72, BMP & BTR (nonUS)	Range 1	Range 3

The Signature Challenge training stressed short drills. The first six sets consisted of 16 images, all eight aspects of two vehicles. Sets 7 and 8 had 24 images, all 8 aspects of 3 vehicles. Each vehicle was exposed for a maximum of ten seconds. The criterion for passing (85%) was no more than two errors for Sets 1 through 6, and no more than three errors for Sets 7 and 8. At the end of each trial, feedback was provided on whether the vehicle selected was correct or incorrect. Total percent correct was provided at the end of each session.

There was no "culminating" vehicle set that included all images, as in Experiments 1 and 2. However, sets 7 and 8 grouped the vehicles differently to give soldiers the opportunity to eliminate possible cross-confusions. These two sets, US and nonUS vehicles, not only separated the friendly from likely threat vehicles, but they also separated vehicles along thermal imagery dimensions. The nonUS vehicles all presented low-profile images, with relatively low thermal signatures and front views often similar to rear views. On the other hand, the thermal cues displayed by the US vehicles were more distinct.

Testing for transfer. Two tests were given: Signature Challenge and Vehicle ID. For each test, 24 images tested soldiers' ability to transfer vehicle identification skills acquired at one range to another distance. Soldiers who trained on near images were tested on far images. Soldiers who trained on far images were tested on near images.

All transfer images were oblique views. The oblique aspects were selected because the first two experiments had shown vehicle identification to be of intermediate difficulty with these views, easiest with flanks, and hardest with the front and rear. Previous research, using fewer than eight aspects, supported these findings. With photopic displays and models, fronts were harder than front obliques and flanks (Haverland & Maxey, 1978; Warnick & Smith, 1989). With thermal images, O'Kane et al. (1997) found the front to be more difficult than a rear oblique and a front oblique view.

Within each range band on the transfer test, the number of images was divided evenly between the closer and the farther range. To reduce test length, vehicles and aspects were counterbalanced within range bands. In addition, the aspect-range combinations were reversed for Signature Challenge and Vehicle ID. An additional 12 images of front and rear views were presented as no transfer conditions. Each test had a total of 36 images. Details on the transfer tests are in Table 3-4.

Table 3-4
Experiment 3: Vehicle ID and Signature Challenge Transfer Tests

Test	Practice Condition	Transfer from Practice		No Transfer from Practice	
		Rt Ft & Lf R Oblique	Lf Ft & Rt R Oblique	Front	Rear
Vehicle ID	Near	Range 2	Range 3	Close-up	Range 1
	Far	Close-up	Range 1	Range 2	Range 3
Signature Challenge	Near	Range 3	Close-up	Range 1	Close-up
	Far	Range 1	Range 2	Range 3	Range 2

Note. Ranges for testing a specific aspect were reversed in Vehicle ID and Signature Challenge.

Ten seconds was allowed for a response in the Signature Challenge test. The Vehicle ID test was not timed; both identification and aspect responses were required. No feedback was given to the soldiers on either transfer test. Half took Vehicle ID first; half, Signature Challenge first.

Pretests and posttests. The pretests and posttests were identical to those in Experiments 1 and 2. As before, visible and FLIR versions of each were given. However, it is important to note that the vehicles trained in this experiment differed slightly from the previous experiments. The FLIR training covered the M1, T72, M2, BMP, LAV and BTR. It did not cover the M113, HMMWV, ZSU, and M60.

Measures of Performance

The pretest, posttest, Vehicle ID exercise and test, and Signature Challenge exercises and test each required soldiers to identify the vehicle displayed. Vehicle images were presented randomly within all exercises and tests. For the pretest and posttest, soldiers had to select their response from a list of all the 14 vehicles within CVI*Plus*. For Signature Challenge training,

only the vehicles being displayed were possible responses. Depending on the particular Signature Challenge vehicle set, this involved a choice of two or three vehicles. For the Signature Challenge test, the soldiers had to select from all six vehicles trained. For all Vehicle ID exercises, the buttons for all 14 vehicles in CVI*Plus* were automatically displayed as options. The soldiers were told that only the six vehicles being trained would be presented and were given a list of these vehicles to make the training and test conditions similar to Signature Challenge.

The Vehicle ID exercises and tests required soldiers to identify the aspect angle as well. The other criterion measure was response time. All Signature Challenge exercises were timed, with a maximum of ten seconds allowed. Actual response time, however, was recorded for all tests and exercises. The number of sessions required to reach criterion on each Signature Challenge vehicle set was also tabulated.

The paper-pencil test on engine and vehicle location is in Appendix D. Following the posttest, soldiers were given a survey on their reactions to the CVI*Plus* program (Appendix E).

Experiment 3: Results

Training and Testing Time

The times for each major training and testing event are shown as a function of the Near and Far practice conditions in Figure 3-1. The two groups are divided by the range bands of the images presented during training, that is, the Vehicle ID and Signature Challenge practice exercises. Group instructional time was 30 minutes.

Clearly, the soldiers in the Far practice condition required more time on both the Vehicle ID and Signature Challenge exercises than those in the Near condition. Total average time for the Far condition was 20 minutes more than the Near condition. The difference in the two conditions resulted primarily from the increased training times for the Far condition, both on Vehicle ID and Signature Challenge. In comparing the slowest and fastest individual times within each condition, the difference was 50 minutes for the Near condition and 70 minutes for the Far condition. Of further interest is that the mean for the Near condition was about the same as the time for the fastest individual in the Far condition. These findings reflect the greater difficulty of images at the farther ranges. Table D-1 provides further information on the times.

Engine and Exhaust Location Test

The engine and exhaust location scores showed that after the group instruction, soldiers typically knew engine location (e.g., in the rear or front). The mean score for the six vehicles was 86% (Table D-2). The mean score for whether the engine was centered or was on the right or left was 70%. The exhaust location score was 79%. Comparisons of the near and far instruction groups showed no significant differences on total engine score and the exhaust location score. The M1 consistently received the highest scores; the BMP, the lowest scores.

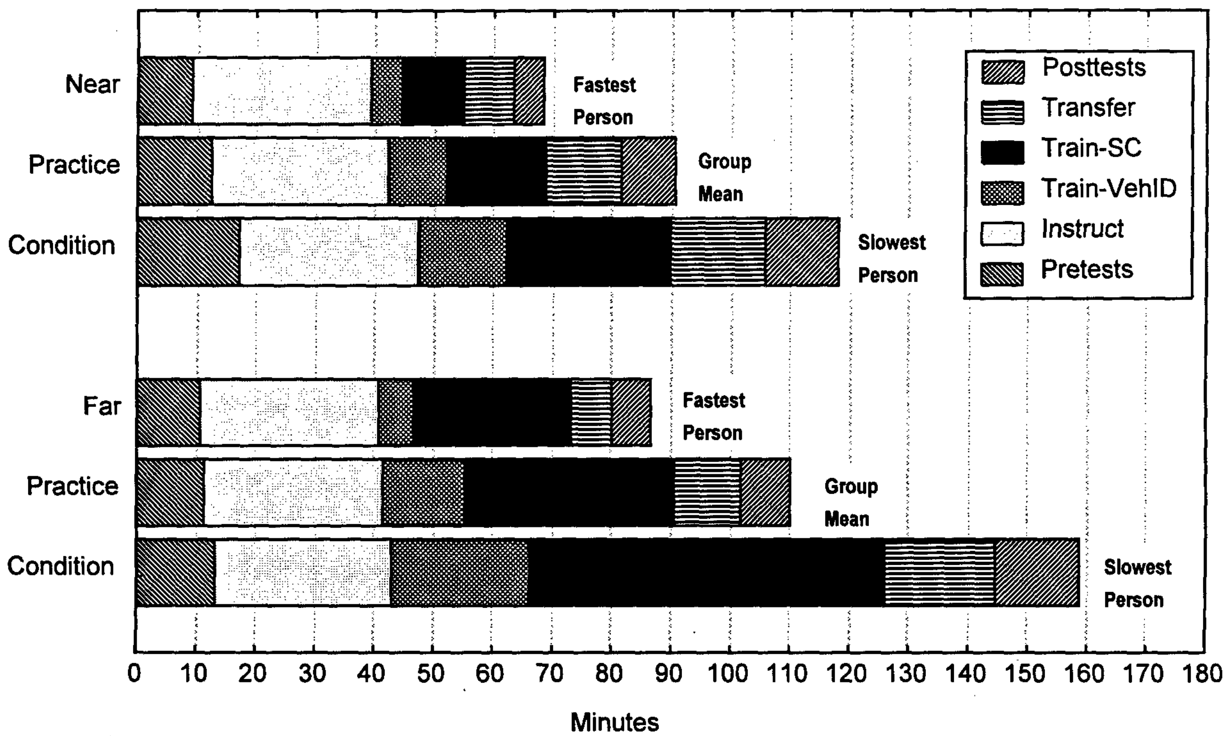

Figure 3-1. Experiment 3: Training and testing times for the Near and Far practice conditions (group means and fastest and slowest individuals).

Training

Vehicle ID training exercises. Vehicle ID was the first training exercise. Unfortunately, the computer used for nine soldiers had a software problem that mixed daytime visible images with FLIR imagery in this module. Consequently, the analysis of the Vehicle ID training data was based on 58 of the 67 soldiers. The nine soldiers were distributed equally across the four experimental conditions.

Analyses were conducted on vehicle identification scores, aspect scores, and response times. The results are presented in that order.

For <u>vehicle identification scores</u>, a 2 x 2 x 6 x 4 (instruction by practice by vehicle by aspect) ANOVA with repeated measures on the last two factors was conducted. Main effects occurred for practice, vehicle, and aspect. There was also a two-way interaction between instruction and vehicle, a two-way interaction between vehicle and aspect, and a three-way interaction among practice, vehicle, and aspect. Complete data on these effects are in Table D-3.

The main effects on Vehicle ID showed that identification scores were higher for the near (78%) than for the far images (65%); higher for the M1, M2, and T72 (81%) than the LAV,

BMP, and BTR (62%); and higher for the flanks (83%) than the front and rear (61%) aspects (Table D-3). The interaction between the instruction and vehicle factors showed no differences between the two instruction groups for the T72 and M2, higher scores for the far instruction group on the M1 and BMP, and higher scores for the near instruction group on the LAV and BTR (Table D-3).

The strong main effect for aspect was attenuated by the interaction between the aspect and vehicle factors. As indicated in the dot plot in Figure 3-2, relatively small differences appeared among the six vehicles for both flanks (at least 75% correct), except for the low score for the right flank of the BMP (50% correct). Greater variability among vehicles occurred for the front and rear aspects. The M1, M2, and T72 had relatively similar and high identification scores from both front and rear aspects (67% to 83%). The BMP scores for the front and rear were also similar, although they were lower (45%). However, for the LAV, the front produced higher identification scores than the rear (60% vs 27%). For the BTR, the rear produced higher identification scores than the front (60% vs 38%).

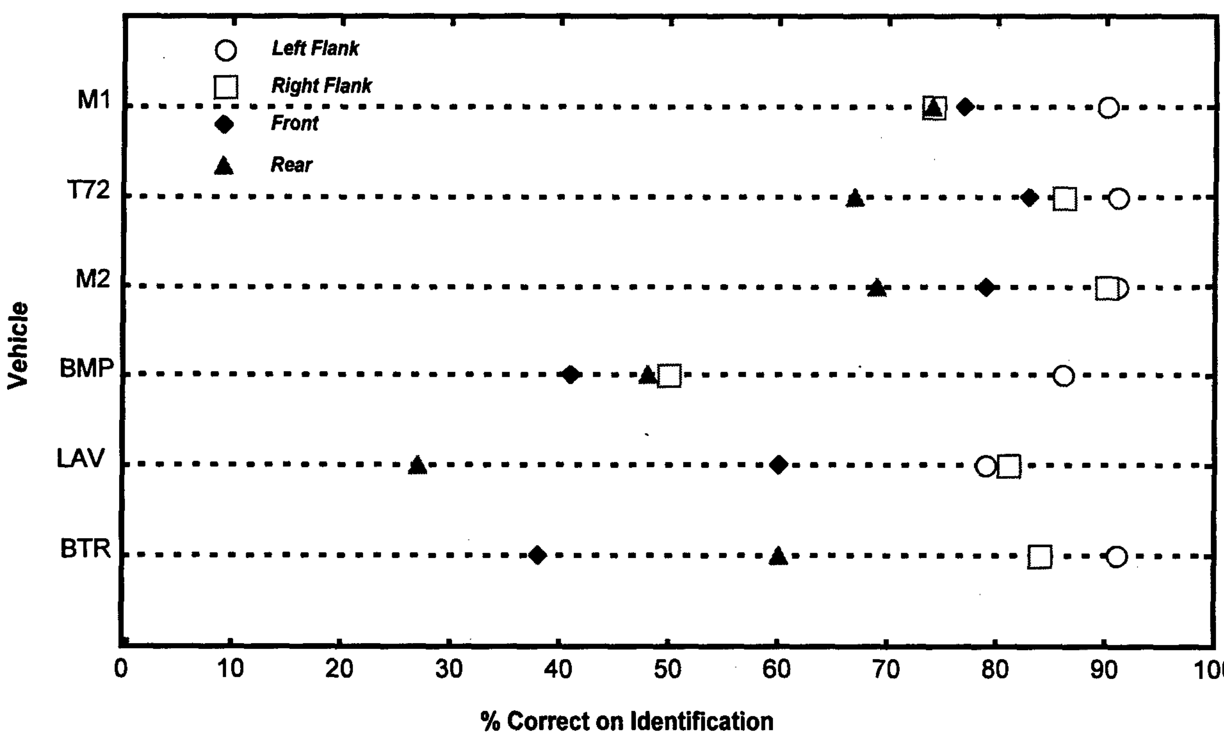

Figure 3-2. Experiment 3: Identification scores on Vehicle ID training exercises as a function of vehicle and aspect presented.

Which vehicles were confused with each other during this initial training session? The vehicle confusion matrix is in Table 3-5. The results are shown by range band. The diagonal cells present percent correct scores; the off-diagonal cells, vehicle confusions. For example,

when the BMP was shown at the near range, 65% of the time soldiers identified it correctly; 15% of the time they incorrectly identified it as a T72. When the BMP was shown at the far range, 48% of the time soldiers identified it correctly. But 25% of the time they incorrectly identified it as a M2 and 16% of the time as a T72.

Table 3-5

Experiment 3: Vehicle Confusion Matrix for Vehicle ID Training

Vehicle and Range Band Displayed	Vehicle Response					
	M1	T72	M2	BMP	LAV	BTR
M1 Near	**87**	13 (Rear) (Right)	0	0	0	0
Far	**71**	22 (Front) (Right)	1	3	0	3
T72 Near	8	**86**	0	3	0	0
Far	12 (Right)	**78**	1	6 (Rear)	1	1
M2 Near	3	0	**91**	3	2	1
Far	3	7	**75**	6	4	3
BMP Near	4	15 (Front) (Rear)	7	**65**	7	3
Far	4	16 (Rear)	25(Front) (Rear) (Right)	**48**	4	1
LAV Near	1	5	16 (Rear)	3	**66**	9
Far	3	6	5	4	**59**	19 (Front) (Left) (Right)
BTR Near	3	3	0	11 (Front (Rear)	8	**76**
Far	12 (Rear)	4	2	10 (Front)	10 Right)	**61**

Note. Percentages for each vehicle for each condition (range) are based on 116 trials (4 aspects, 29 soldiers). Diagonal cells cite percent correct identifications; off-diagonal cells are error rates. When an aspect resulted in an error rate of at least 15%, it is cited. For example, when the rear of the M1 was shown at the near range, at least 15% of the time it was confused with a T72. The row percentages may not sum to 100, as about 1% of the time soldiers selected vehicles not presented.

The aspect angles that created vehicle confusions are also cited in Table 3-5. When a particular aspect resulted in a high confusion with another vehicle, that aspect is cited. High confusion was defined as an error rate of at least 15%. For example, consider the M1. As cited in Table 3-5, when the rear of the M1 was shown at the near range band, at least 15% of the time soldiers confused it with a T72. At the same range band, when the right flank of the M1 was presented, it was also likely to be misidentified as the T72.

Some errors are easily explained given the thermal signature of the vehicles. For example, the right flanks of the two tanks tended to be confused with each other, but not the left flanks. The exhaust of the T72 on its left flank is a distinguishing feature. When this cue is not observable, the two tanks appear similar, particularly at longer ranges. On the other hand, some errors appear simply to be confusions that occur during the normal process of learning to discriminate vehicles. For example, when the rear of the BMP was shown, it was sometimes misidentified as the T72. With both vehicles, the exhaust can be seen from the rear. However, for the BMP the exhaust is on the right; for the T72, the exhaust is on the left.

The Vehicle ID training exercise also required that soldiers determine which aspect was presented. Aspects were easier to determine than vehicles, 85% correct (SD = 10) for aspect scores versus 72% correct (SD = 15) for vehicle identification scores, $t(57) = 7.35, p < .0000$.

A repeated measures analysis was also conducted on the aspect scores. This analysis paralleled that conducted on the identification scores, that is, it was a 2 x 2 x 6 x 4 (instruction by practice by vehicle by aspect) ANOVA with repeated measures on the last two factors. Main effects occurred for practice, vehicle, and aspect. Two-way interactions occurred between practice and vehicle, instruction and aspect, practice and aspect, and vehicle and aspect. The only three-way interaction was between practice, vehicle, and aspect. Complete data on these effects are in Table D-4.

The main effects showed that determining the cardinal aspect of a vehicle was easier at near than far ranges (90% versus 78%). The left flank was the easiest (98%), then the right flank (87%), and the front and rear were the hardest (77%). Aspect scores for the six vehicles were ordered from high to low as follows: M1, T72, M2, BMP, LAV, and BTR. This vehicle order held for both far and near practice conditions. However, aspect scores declined steadily in accordance with this vehicle order at the far practice condition (from 97% to 70%), while the decline was less sharp under the near practice condition (from 98% to 84%, Figure 3-3). Thus the difficult aspects became even more difficult to determine at the longer ranges.

All other interactions involved the aspect displayed. The practice by aspect interaction showed that the left flank was easily determined at both ranges, while the far range created more difficulty for the other aspects. The aspect confusion matrix in Table D-5 shows that fronts and rears were confused with each other at both range bands, and the right flank confused with the left flank at the far range band.

The main conclusion to be drawn from the other interactions is that the ease with which the cardinal aspect of a vehicle can be determined during initial training varies with the vehicle as well as the range. For example, the vehicle by aspect interaction in Figure D-1 shows that aspect scores were consistently high for all vehicles when the left flank was displayed; there was greater disparity among vehicles for the right flank. The front and rear aspects produced the greatest differences among vehicles. High scores (> 90%) on all aspects occurred only for the M1 tank. These results parallel in many instances the results on vehicle identification scores depicted previously in Figure 3-2, where high identification scores occurred for most vehicles

when the left flank was presented. But when the front and rear were presented, the ease with which vehicles were identified varied considerably.

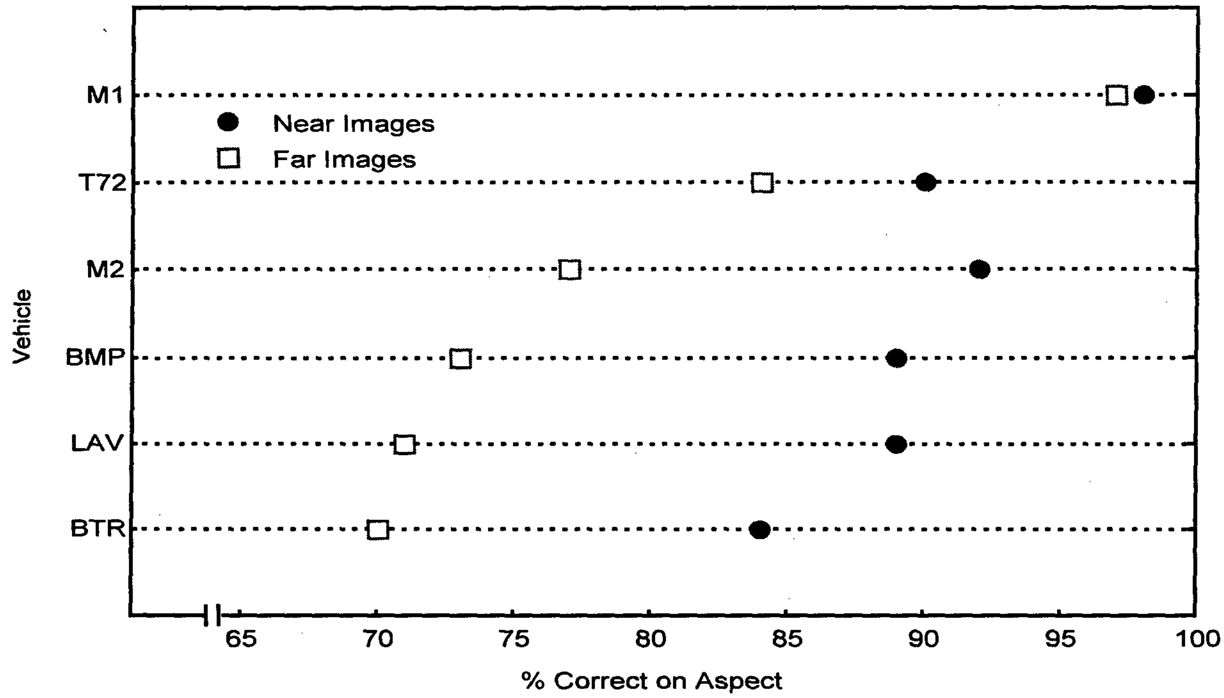

Figure 3-3. Experiment 3: Aspect scores on Vehicle ID training exercises as a function of vehicle and far and near practice conditions.

<u>Time</u> to identify the vehicle and indicate its aspect was also examined. On average, the response time per trial for soldiers presented with the near images was 22.83 sec (*SD* = 24.80). Those presented with the far images averaged 32.05 sec (SD = 36.18). Response times were 1.4 times longer for the far images. The large standard deviations reflect the large variability in response times for individual soldiers in this initial training.

For both conditions, responses were quickest to the M1 (Table 3-6). Longest response time for the near practice condition was for the BMP, with soldiers taking 1.5 times as long to respond to the BMP as to the M1. For the far practice condition, the longest response time was to the LAV, with soldiers taking 1.8 times as long to respond to it as the M1.

Response times were also examined as a function of the four cardinal aspects. For both the far and near range bands, the soldiers' response times were about 1.7 times longer when the front and rear aspects were displayed than when the flanks were displayed (see Table 3-6).

Signature Challenge training exercises. With the exception of five soldiers in the far practice group, all soldiers reached the criterion of 85% correct vehicle identifications on each of

the Signature Challenge vehicle sets. Because of time constraints, these five soldiers did not reach criterion on the eighth and final set; they did reach criterion on all other sets. On this final set they achieved scores of at least 79% correct. For analytic purposes, no further distinction is made between these soldiers and the others; they are "treated" as if they did achieve criterion on the final set. Analyses were based on the complete sample of 67 soldiers.

Table 3-6
Experiment 3: Mean Response Time (sec) per Trial by Vehicle and Aspect in Vehicle ID Training Exercises

Vehicle and Aspect		Near Practice *M* (*SD*)	Far Practice *M* (*SD*)
Vehicle[a]	M1	18.37 (25.95)	25.18 (37.23)
	M2	19.17 (22.21)	29.64 (31.93)
	T72	22.06 (22.32)	30.09 (39.93)
	BTR	22.66 (21.41)	28.38 (31.89)
	LAV	25.34 (26.87)	45.24 (43.07)
	BMP	29.38 (28.07)	33.78 (28.85)
Aspect[b]	Right Flank	17.31 (17.88)	25.02 (29.98)
	Left Flank	16.55 (21.12)	21.28 (23.04)
	Rear	28.88 (29.79)	39.51 (35.66)
	Front	28.58 (26.03)	42.39 (47.20)

[a] Means based on 116 trials per vehicle (29 soldiers, 4 aspects).
[b] Means based on 174 trials per aspect (29 soldiers, 6 vehicles).

The number of sessions required to reach criterion on each set for the near and far practice conditions is displayed in Figure 3-4. In general, the soldiers in the far practice condition required twice as many sessions as those in the near condition. The relative ease of identifying vehicles at the near range band is reflected in the overall mean of 1.23 sessions (*SD* = .33) for the near condition compared to 2.23 sessions (*SD* = .86) for the far condition. No one in the far practice condition averaged one session per vehicle set, while 12 (36%) did in the near practice condition. In contrast, 16 (47%) of the soldiers in the far practice condition required an average of two or more sessions, while only one (3%) did so in the near practice condition. A Chi-square comparing the number of sessions for the two practice groups was significant, $\chi^2(2) = 25.33, p < .0000$.

Of interest was the number of sessions to reach criterion for the last two sets. The next to last set (#7) had all the US vehicles. The last set (#8) had all the nonUS vehicles. The relative ease with which the soldiers discriminated the US vehicles (see Figure 3-4) stands in contrast to the greater difficulty in discriminating the nonUS vehicles, particularly at the far ranges (M = 2.91, SD = 1.83) versus the near ranges (M = 1.54, SD = 1.17). It is hypothesized that the greater similarity in the thermal images of the nonUS vehicles (low profile, many parts appearing relatively cool, similar fronts and rears) contributed to this result. In contrast, the US vehicles

are more distinct. This finding reinforces the need to give soldiers exercises requiring them to discriminate similar vehicles, and to provide several vehicle combinations during training.

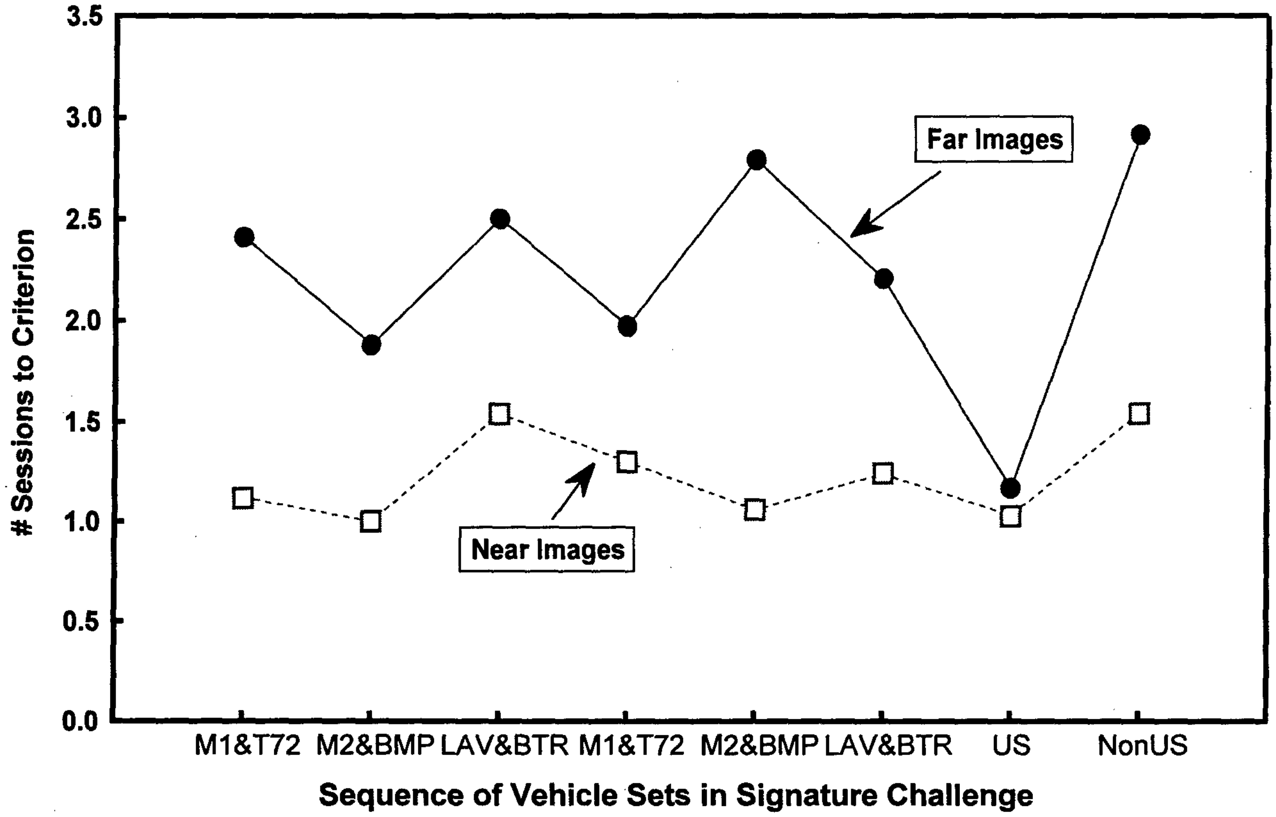

Figure 3-4. Experiment 3: Number of sessions to criterion for far and near practice conditions.

Table D-6 shows the mean identification scores for each vehicle during the first session of each Signature Challenge training set. The first session was used as that was the only session available on all soldiers. Overall, the first-session scores were 10% higher for the near than the far range, 92% and 81% respectively. Vehicles that were more difficult to identify initially continued to be the more difficult to identify as the range increased on later sets. For example, the M1 was easier to identify than the T72; the M2, easier to identify than the BMP. Time to respond showed a similar pattern (see Table D-6). Soldiers answered more quickly across vehicle sets, regardless of range, when they knew the vehicle.

Consistent relationships also emerged between identification scores and response time for the near and far images when vehicle aspect was considered. Based on the first session from all the vehicle sets, identification scores and response time as a function of the eight aspects were inversely correlated. The correlation between the means in Table 3-7 for the near range was -.91, $p < .05$; for the far range, it was -.87, $p < .05$. Thus, the aspects associated with the highest scores were those with the fastest response times. Aspects that created more difficulty in

determining the vehicle had slower response times. In addition, the correlation between identification scores on the near and far ranges was high ($r = .87, p < .05$), as was the correlation between response times on the near and far ranges ($r = .91, p <.05$). These relationships support prior research findings on the mediating role of vehicle aspect on soldier ability to identify vehicles.

Table 3-7

Experiment 3: Mean Identification Scores and Response Times by Aspect Angle on the First Session of all Eight Signature Challenge Vehicle Sets

Aspect Presented	Identification Score (% correct)		Time to Respond (sec)	
	Near Range	Far Range	Near Range	Far Range
Right Flank	92	87	1.50	1.79
Left Flank	92	88	1.51	1.69
Right Rear Oblique	90	80	1.83	2.30
Left Rear Oblique	92	82	1.76	2.01
Right Front Oblique	92	78	1.83	2.34
Left Front Oblique	89	77	1.86	2.30
Rear	86	76	2.02	2.33
Front	85	66	2.18	2.44
Mean - All Aspects	89	79	1.81	2.15

Note. Means based on 594 trials for near range; 612 trials for far range.

Transfer of Skill

For both the Signature Challenge and Vehicle ID tests, results on the front and rear aspects are presented separately from results on the oblique aspects. The front and rear aspects were the no transfer conditions, as they were presented at the same range band used during the training exercises. The oblique aspects were used to test transfer of skill to the other range band. We were aware of the confounding between aspect and the transfer and no transfer conditions (i.e., front and rear harder than oblique views), and therefore that these two conditions were not directly comparable. For both tests, all six vehicles were presented in a single session, and there was no feedback either during or after the tests. Vehicle ID, requiring both vehicle identification and aspect responses, was not timed. Signature Challenge, requiring only vehicle identification responses, was timed, with vehicles exposed for 10 seconds.

Test imagery trained at the same range (no transfer). A direct assessment of the effect of the training exercises on identifying vehicles from the front and rear aspects was obtained by comparing the Vehicle ID training and test scores. This comparison was made only for those soldiers ($n = 58$) who had valid Vehicle ID training imagery. Identification scores on Vehicle ID for these two aspects improved significantly at both range bands on the test, from 60% to 75% correct, $F(1, 55) = 33.73, p < .0000$. For the near ranges, scores increased from 69% to 79%. For the far ranges, scores increased from 52% to 69% correct. The difficulty in identifying

vehicles from the front and rear at the far range band was evident, as the transfer score for the far ranges was the same as the initial training score for the near ranges.

Also of interest was the general progress in scores on the no transfer aspects throughout training and testing. What were the effects of Signature Challenge training? Would identification scores be the same as those achieved at the end of Signature Challenge training, despite differences in the test conditions and the test imagery pool? The box and whisker plots in Figure 3-5 provide a picture of identification scores for the front and rear aspects over the course of the training and testing. Each plot is based on 12 scores; the mean percent correct for the front and rear of each vehicle (refer to Table D-7). The box plots show the minimum and maximum values for these means, the middle 50%, and the median.

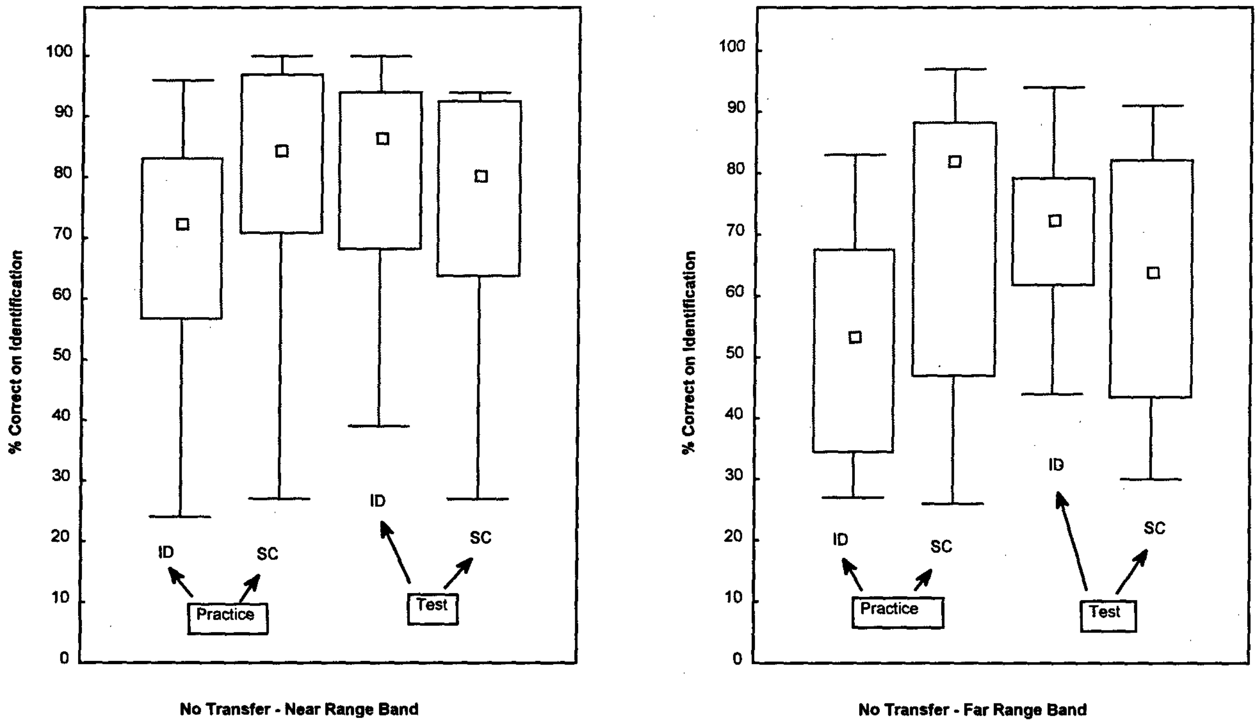

Figure 3-5. Experiment 3: Box plots of identification scores for front and rear aspects (no transfer) during training exercises and testing.

For each range band in Figure 3-5, four box and whisker plots are shown. Vehicle ID training exercise data are presented first. The second box plot reflects the last two Signature Challenge training sets (#7 with the US vehicles and #8 with the nonUS vehicles). First session results on these two sets are presented. These Signature Challenge data are an underestimate of training proficiency, as everyone had to reach 85% correct before going to the next set; everyone did not reach criterion on the first session. The last two box plots portray results from the two

transfer tests. The order of taking these two tests was counterbalanced in the experiment, so their sequence in Figure 3-5 is arbitrary.

Three trends are evident from Figure 3-5. First, proficiency was maintained better with the near than the far imagery, despite being trained to the same 85% criterion in Signature Challenge. Second, for each range band, scores improved during training; Signature Challenge training scores were higher than the initial Vehicle ID scores. Third, for each range band, scores on the Vehicle ID test tended to be higher than scores on the Signature Challenge test.

Analyses were made of the no transfer test scores (front and rear aspects) to examine the effect of the range band of the imagery and the test format. A 2 x 2 x 2 x 6 (instruction x practice x test x vehicle) ANOVA with repeated measures on the last two factors was conducted. Main effects occurred for factors of practice (range of the images), test format (Veh ID or SC), and vehicle. There was a two-way interaction between practice and vehicle, a three-way interaction with instruction, practice, and vehicle, a three-way interaction with practice band, test and vehicle, and a four-way interaction involving all factors. Complete data on these effects are in Table D-8.

Main effects showed higher identification scores for the near images than for the far images (77% versus 66%). Scores were higher for the M1, M2, and T72 (82%) than for the LAV, BMP and BTR (62%). They were higher for the Vehicle ID test (75%) than for the Signature Challenge test (69%). Main effects for practice paralleled those during Vehicle ID training (reference Table D-3). The tendency for Vehicle ID scores to be higher than SC scores probably reflects the non-timed conditions of the ID test and the opportunity to change an answer before proceeding to the next vehicle.

The interaction between vehicle and the practice condition showed that, across both tests, the near images were easier than far images for all vehicles except the LAV (Figure 3-6). As also shown in Figure 3-6, the difference between the near and far range bands was greatest for the M2 (20% difference) and least for the T72 and BMP (10% difference).

The variability in scores appears to reflect the difficulty of identifying vehicles when the front and rear are shown, as well as interference from the transfer images. The front and rear aspects were the most difficult. If the flanks had been presented as no transfer images, the scores would probably have been less variable and the differences between the near and far imagery smaller.

Which vehicles were confused when the front and rear views were presented? Were they the same no matter which test format? Vehicle confusion matrices are in Table 3-8. The diagonal cells, the percent correct identifications, are in bold print. Within each cell, the first line presents results with near imagery from the Vehicle ID test; the second, the Signature Challenge test. The third line is far imagery from the Vehicle ID test; the fourth, the Signature Challenge test. As with the other confusion matrices, when a particular aspect resulted in confusion rate of at least 15% with another vehicle, that aspect is noted.

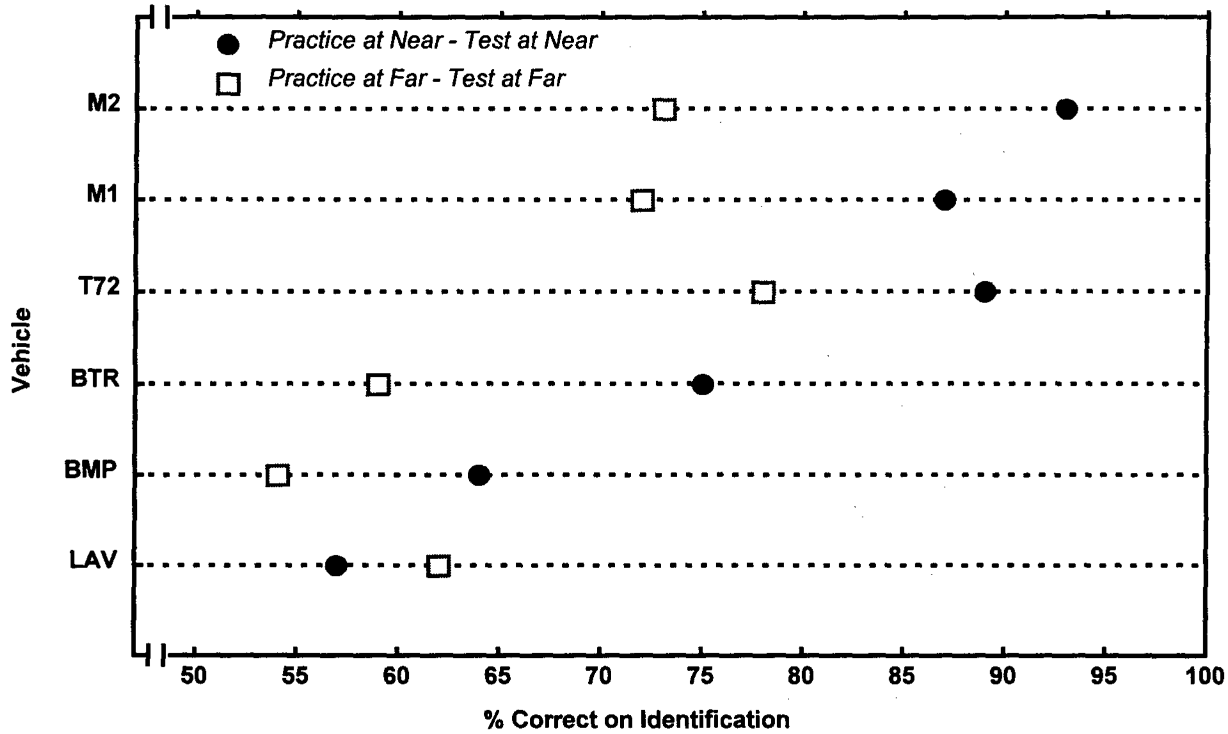

Figure 3-6. Experiment 3: Identification scores on tests for front and rear aspects (no transfer) as a function of vehicle and practice-to-test range band.

Several vehicle confusions stand out regardless of test format. The BMP was confused with the T72 and the M2 when seen from the front. When seen from the front, the BTR was consistently confused with the BMP, and was confused with the T72 at far ranges. When the LAV was viewed from the rear, soldiers often thought it was the M2 or the BTR instead. It is important to note that it is often difficult to distinguish the thermal signature of wheels from tracks from the front and rear aspects. Typically, the confusions were not symmetric, that is, when the front of the BMP was presented it was confused with the M2, but when the front of the M2 was presented it was not confused with the BMP.

Test imagery trained at a different range (transfer). The primary question addressed by the transfer imagery was whether vehicle identification skills transferred to a different range band. Soldiers trained on near imagery were tested on far imagery; those trained on far imagery were tested on near. It was anticipated that going from far to near would be easier than the reverse. The box and whisker plots in Figure 3-7 provide a general picture of the transfer that occurred. Each plot is based on 24 scores; the mean percent correct for each of the oblique angles for each vehicle (Table D-9). No data are presented on Vehicle ID training, as oblique angles were not in that training. Signature Challenge training scores are from the first session of the last two exercises and are based on all soldiers.

Table 3-8

Experiment 3. Vehicle Confusion Matrices for Front and Rear Aspects (no transfer) on the Vehicle ID and Signature Challenge Tests

Vehicle and Range Band Displayed	Vehicle Response					
	M1	T72	M2	BMP	LAV	BTR
M1 Near ID	**85**	11 (rear)	3	0	0	0
Near SC	**91**	8	0	0	0	0
Far ID	**84**	12 (front)	3	1	0	0
Far SC	**63**	10	16 (front)	8 (front)	2	2
T72 Near ID	6	**94**	0	0	0	0
Near SC	5	**85**	2	6	3	0
Far ID	7	**82**	1	6	1	1
Far SC	6	**75**	2	11 (front)	0	6
M2 Near ID	0	2	**92**	5	0	2
Near SC	0	0	**94**	5	2	0
Far ID	1	1	**76**	10	3	7
Far SC	2	3	**71**	15 (rear)	6	3
BMP Near ID	0	15 (front)	2	**68**	11 (rear)	5
Near SC	0	18 (front)	6	**62**	8	6
Far ID	3	15 (front)	13 (front)	**54**	9	6
Far SC	6	5	18 (front)	**54**	5	12 (rear)
LAV Near ID	2	2	18 (rear)	10	**62**	6
Near SC	0	3	34 (rear)	6	**52**	5
Far ID	1	3	10	9	**63**	13 (rear)
Far SC	0	2	20 (rear)	5	**64**	11 (rear)
BTR Near ID	5	2	0	14 (front)	5	**76**
Near SC	3	0	0	14 (front)	8	**76**
Far ID	6	15 (front)	3	13 (front)	4	**59**
Far SC	6	12 (front)	5	15 (front)	2	**61**

Note. An aspect is cited only when that view of the displayed vehicle resulted in misidentifications at least 15% of the time (e.g., when the front of the BTR was displayed, it was likely to be misidentified as a BMP). Vehicle % based on 66 trials for near ranges; 68 trials for far ranges.

Two trends are indicated in Figure 3-7. First, as expected, transfer was greater when going from far to near imagery, than vice versa. Soldiers who trained on the far images did well on the near images. Their transfer scores on near images were comparable to their final training scores on the far images (within 5% points). The picture was different for soldiers who were trained on the near images and tested on far images. On average, their transfer scores on far were at least 20% points lower than their final training scores on the near images. The second trend evident in the box plots is that when soldiers were trained on far, there was little variability upon transfer to the near imagery. Yet when soldiers were trained on near, there was substantial variability upon transfer to far imagery.

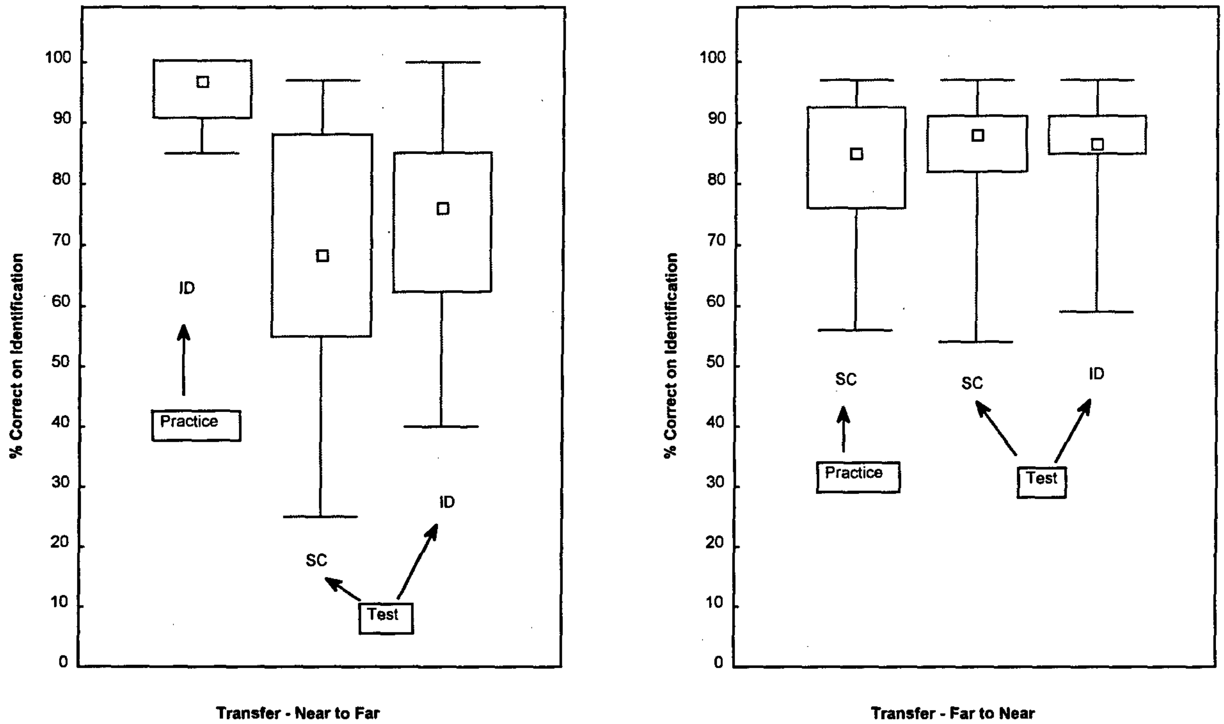

Figure 3-7. Experiment 3: Box plots of identification scores for oblique aspects (transfer) during training exercises and testing.

For the transfer aspects, a 2 x 2 x 2 x 6 (instruction x practice of the imagery x test x vehicle) ANOVA with repeated measures on the last two factors was conducted on the vehicle identification scores. Main effects occurred for the factors of practice, test, and vehicle. There was a two-way interaction between practice and vehicle, a two-way interaction between test and vehicle, and a three-way interaction with instruction, practice and vehicle. Data on these effects are in Table D-10.

The main effects showed that identification scores were higher for the near than the far images (85% versus 71%). Scores on the Vehicle ID test were higher than those on the Signature Challenge test (80% versus 76%). Of the six vehicles, the M1 and M2 had the highest scores (84% and 89% respectively); the BMP, the lowest (64%, see Table D-10). Several significant interactions moderated these main effects. These interactions are discussed and illustrated in Appendix D (Figures D-2, D-3, and D-4).

<u>Vehicle confusion</u> matrices for the oblique aspects are in Table 3-9. The table format is the same as that for the front and rear aspects in Table 3-8.

Table 3-9

Experiment 3: Vehicle Confusion Matrices for the Oblique Aspects (transfer) on the Vehicle ID and Signature Challenge Tests

Vehicle and Range Band Displayed	Vehicle Response					
	M1	T72	M2	BMP	LAV	BTR
M1 Near ID	**85**	10 (rt ft)	2	1	0	1
Near SD	**88**	5	5	2	0	1
Far ID	**78**	8 (rt ft)	8	5 (lf ft)	1	0
Far SC	**83**	12 (lf ft)	2	2	0	1
T72 Near ID	6	**90**	0	2	0	1
Near SC	15 (lf ft) (rt ft)	**78**	1	5	1	1
Far ID	11 (rt rear)	**75**	0	13 (rt ft)	0	0
Far SC	18 (lf rear) (rt ft) (rt rear)	**60**	5	11 (rt ft)	1	3
M2 Near ID	4	1	**93**	2	0	0
Near SC	3	1	**92**	2	2	0
Far ID	5	1	**83**	8	1	1
Far SC	0	2	**91**	5	0	2
BMP Near ID	2	6	9 (rt rear)	**76**	4	2
Near SC	2	8 (rt ft)	10 (rt rear)	**76**	4	2
Far ID	1	10 (lf rear)	17 (lf ft) (rt rear)	**61**	5	6
Far SC	6	11 (lf ft) (rt ft)	27 (lf rear) (rt rear) (rt ft)	**47**	6	3
LAV Near ID	1	0	2	1	**89**	7
Near SC	0	0	1	0	**87**	12 (lf ft)
Far ID	2	1	10 (rt rear)	5	**70**	11
Far SC	5	2	12 (lf ft) (rt ft)	6	**62**	12 (lf ft) (lf rear)
BTR Near ID	1	0	0	4	8	**87**
Near SC	0	1	1	4	11 (lf ft)	**83**
Far ID	1	4	3	8	10	**73**
Far SC	2	6 (rt ft)	4	11 (rt ft)	6	**71**

Note. Some rows may not sum to 100% because of rounding. An aspect is cited when that view of the displayed vehicle resulted in misidentifications at least 15% of the time. Vehicle percentages based on 132 trials for near ranges; 136 trials for far ranges.

The oblique views created confusions between the two tanks, the T72 and the M1. The BMP was most often confused with the M2, the other tracked APC, when a rear oblique was shown. The BMP tended to be confused with the T72 when a front oblique was shown. The LAV was confused with both the M2 and the BTR. The BTR was the only other wheeled APC in the test pool. The LAV, although wheeled, resembles the M2 in many other respects, as the LAV is modeled after the M2. Of interest is that the soldiers did not incorrectly identify the M2 Bradley (their own vehicle) with other vehicles, regardless of range and type of test.

The Vehicle ID test also provided an opportunity to examine which aspects were confused with each other. Soldiers were not told which aspects were presented, so they could select any of the eight aspects. The aspect confusion matrix is in Table D-11. As during training, the front and rear aspects were confused with each other, more so at the far than the near range band (25% versus 10%). Although the flanks were never presented on the transfer tests, soldiers sometimes confused them with a front oblique view, particularly at the far ranges. Specifically, the left front was likely to be seen as the left flank; the right front, as the right flank. In general, an oblique aspect was only infrequently confused with another oblique aspect.

Overall, response times were consistent with the identification scores. For both tests, soldiers responded faster to the near images than the far images, regardless of the range band they saw during training. Table 3-10 shows these results. The standard deviations were also larger for the far range band in both tests, another indication of the increased difficulty of identifying vehicles at distance.

Table 3-10
Experiment 3: Mean Response Times (sec) on Vehicle ID and Signature Challenge Tests

Practice Condition and Imagery Displayed on Transfer Test	# trials	Vehicle ID Test *M* (*SD*)	Signature Challenge Test *M* (*SD*)
Near Practice			
Test: Front/Rear Aspects at Near	396	11.82 (8.57)	3.31 (2.04)
Test: Oblique Aspects at Far	792	16.52 (13.07)	3.96 (2.33)
Far Practice			
Test: Front/Rear Aspects at Far	408	15.19 (19.83)	4.21 (2.34)
Test: Oblique Aspects at Near	816	11.23 (9.04)	3.13 (1.74)

Pretests and Posttests

As in Experiments 1 and 2, pretest and posttest identification scores were compared for the vehicles included in the FLIR training and those not in FLIR training and whether the image was visible or FLIR. A 2 x 2 x 2 x 2 x 2 (instruction x practice x pre-post x visible-FLIR x train-no train) ANOVA with repeated measures on the last three factors was conducted. The results showed two clusters of effects: one involving the pre-post tests, visible-FLIR, and train-notrain factors; the other involving the practice and instruction factors.

As in Experiments 1 and 2, FLIR training increased identification scores for vehicles in the training, and had a greater impact upon FLIR images than visible images (29% increase for FLIR images vs 15% for visible), although both scores increased (see Table 3-11). On the other hand, both FLIR and visible scores for the vehicles not in the training showed only a slight increase. Figure 3-8 illustrates the three-way interaction. The lowest pretest scores were for the FLIR images of the vehicles in the FLIR training. Posttest scores for the same vehicles, both visible and FLIR images, were very similar.

Table 3-11
Experiment 3: Mean Percentage of Vehicles Identified Correctly on the Pre- and Posttests as a Function of FLIR Training and Type of Imagery

	Visible Imagery			FLIR Imagery		
Vehicle	Pretest	Posttest	Difference: Post - Pre	Pretest	Posttest	Difference: Post - Pre
Vehicles with FLIR Training						
M2/M3	93	97	4	83	94	11
BTR	48	73	25	40	90	50
M1	92	95	3	77	89	12
T72	67	82	15	65	89	24
BMP	62	78	16	48	83	35
LAV	61	85	24	36	73	37
Vehicles without FLIR Training						
M113	97	98	1	95	98	3
HMMWV	95	96	1	89	94	5
M60	65	67	2	51	60	9
ZSU	34	43	9	18	34	16
Trained[a]	70	85	15	58	87	29
Not Trained[a]	73	76	3	63	72	9
Overall[bc]	71	81	10	60	81	21

Note. $N = 67$ soldiers (3 displays per vehicle). Significant main and two-way effects with the Pre-Post, Vis-FLIR and Trn-NoTrn factors.
Pre-Post: $F(1, 63) = 124.33, p < .0000$ Trn-NoTrn: $F(1, 63) = 13.06, p < .0006$
Vis-FLIR: $F(1, 63) = 65.74, p < .0000$ Pre-Post x Trn-NoTrn $F(1, 63) = 61.72, p < .0000$
[a] Pre-Post x Vis-FLIR x Trn-NoTrn $F(1, 63) = 10.82, p < .0016$.
[b] Pre-Post x Vis-FLIR $F(1, 63) = 48.13, p < .0000$
[c] Vehicle means weighed equally in overall mean.

Several interactions involved the practice factor (see Table D-12 for these effects and corresponding means). For FLIR images, scores increased more on the posttest for soldiers who practiced with the near imagery (24%) than those who practiced with the far imagery (14%). Yet scores on the visible images increased the same regardless of the range of the practice imagery.

In addition, for the vehicles that were trained, posttest scores increased more for soldiers who practiced with near images (28%) than for those with the far images (16%). Range of practice imagery had no effect upon the change in scores for the vehicles not included in the training. The results also showed that the substantial increase for the near practice group on FLIR imagery for the vehicles in the training was due primarily to the subgroup that had near imagery during both group instruction and practice exercises. Their scores increased 28%. For the other instruction-practice subgroups, FLIR scores increased an average of 16%. It should be noted that this last interaction could reflect the benefits of having imagery at the same range band for group instruction, practice, and the posttest.

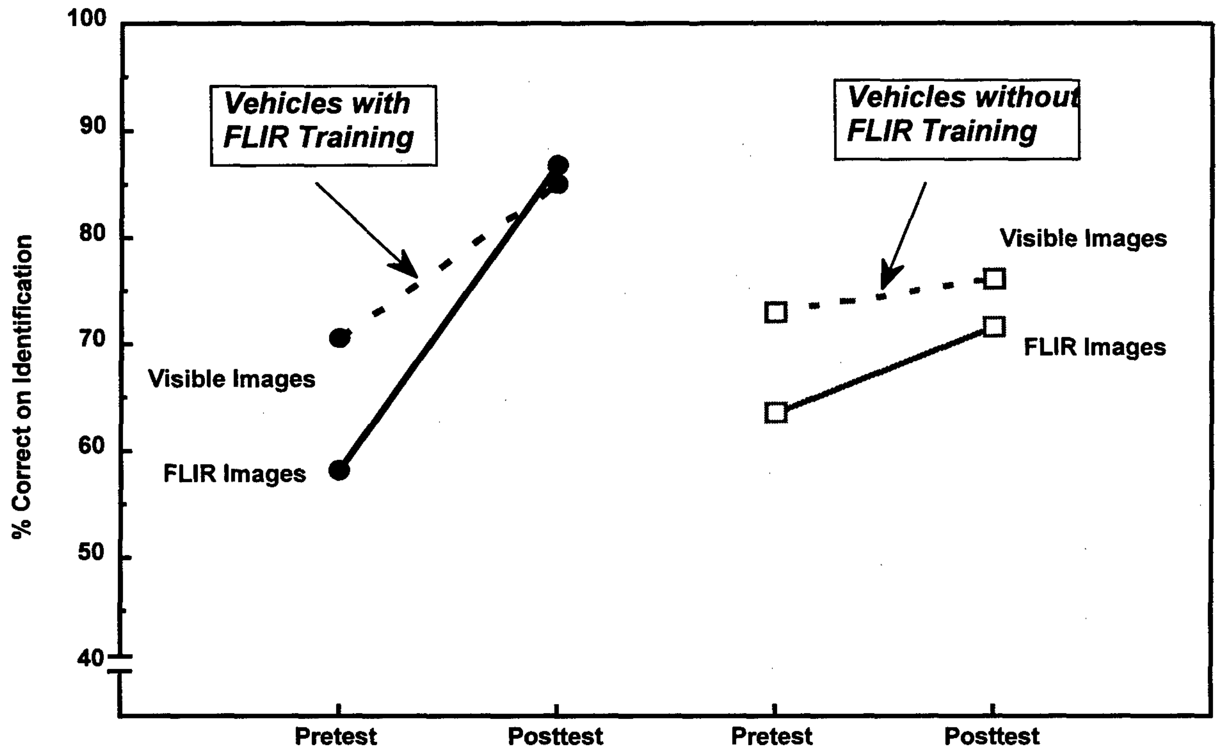

Figure 3-8 Experiment 3: Change in identification scores from pretest to posttest for FLIR and visible images as a function of FLIR training

Vehicle confusion matrices for the pre- and posttests are in Table D-13. Throughout this report, a 15% error rate has been used to describe vehicles likely to be confused with each other. This error rate did not occur for the six vehicles after training, for either the FLIR or the visible imagery, although such error rates did occur before training. The one exception, the LAV, was misidentified as the M2 16% of the time. Of the vehicles not included in the training, the results in Table D-13 show the ZSU remained unfamiliar to the soldiers.

Time to respond was also examined. A 2 x 2 x 2 x 2 (instruct x practice x pre-post x visible-FLIR) ANOVA with repeated measures on the last two factors was conducted. Main effects occurred for the pre-post, and visible-FLIR factors. There was a two-way interaction

between these two factors as well (Table D-14). Response times for the visible imagery were relatively constant from pretest to posttest (8.01 sec and 6.07 sec, respectively). However, response times for the FLIR imagery were more than twice as fast on the posttest than on the pretest, decreasing from 15.28 sec to 6.99 sec. Overall, the response times to the visible images were faster for than the FLIR images. This result parallels the results on time in Experiment 2.

Correlations Among Scores

In general, scores on all tests correlated highly and significantly with each other. The relationship between the test scores and the training scores was not as strong (see Tables D-15 through D-17).

The pretest and posttest scores related significantly to each other. The correlations ranged from .55 to .86 (see Table D-15). This pattern replicated that in Experiments 1 and 2. With the transfer tests, the Vehicle ID and Signature Challenge transfer and no transfer subscores related to each other (r = .67 for transfer; r = .58 for no transfer), indicating that the test format did not change the rank ordering of soldiers greatly (Table D-16). However, there was no significant relationship between the transfer and no transfer subscores, either within or between tests.

Training scores correlated with each other. Vehicle ID training scores exercise correlated negatively with number of sessions to reach criterion on Signature Challenge (r = -.58), indicating that those who achieved the higher scores on the Vehicle ID exercise took fewer sessions in Signature Challenge training. These two training scores were also related to the no transfer subscores (Table D-16), but not to the transfer subscores.

The pretest and posttest scores correlated with scores on the Vehicle ID training exercises; only the posttest scores correlated with number of sessions on the Signature Challenge test (see Table D-17). These Signature Challenge session and posttest correlations were negative, indicating higher posttest scores tended to be associated with fewer sessions to reach criterion.

In summary, individual differences in soldier expertise relative to other soldiers, as assessed by most test scores, did not change dramatically as a result of training. All soldiers became more skilled, but those at the top tended to stay at the top. Training did produce some shifts in the relative positions as evidenced in the lower correlations between training and some test scores. Of interest is that the strength of the relationship to the transfer and no transfer test scores differed for the pretest measures and the training measures. Pretest scores correlated more highly with the transfer test scores (mean r = .65) than the no transfer scores (mean r = .38). Vehicle ID Training scores and Signature Challenge sessions correlated more highly with the no transfer test scores (mean $r = |.57|$) than the transfer scores (mean $r = |.14|$). These findings indicate that when tested on the trained imagery, the best predictors of performance were indices of training progress. Soldiers who did well in training also did well when tested on the same imagery; those soldiers who did less well, scored lower when tested on the same imagery. But

when tested for transfer, the soldier's initial expertise was the better predictor of performance, not training progress.

Experiment 3: Discussion

Answers to the Questions

- Is the amount of training required to identify vehicles at near range the same or different from the amount of training required to identify vehicles at farther ranges?

Twice as much training was needed to learn to identify vehicles at the far as compared to the near range band. Soldiers shown the far imagery required twice as many training sessions on the Signature Challenge exercise to reach the vehicle identification criterion of 85% correct. Total time for the Vehicle ID and Signature Challenge training exercises themselves was twice as long for the far than the near imagery (50 min compared to 26 min). Clearly, the vehicles at distance were more difficult to learn to identify.

- Does skill at identifying vehicles at near ranges transfer to farther ranges?

- Does skill at identifying vehicles at far ranges transfer to nearer ranges?

Soldiers did transfer skills to another distance, when the oblique aspects were presented. As was expected, the easiest transfer was from the far to the near imagery (M = 85% correct). More difficult was the transfer from near to far imagery (M = 71% correct). But ease of transfer varied with the vehicle, despite training that required soldiers to reach the same identification criterion for all vehicles. At the near range, only one vehicle fell below the overall mean of 85%, the BMP. At the far range, there was a greater spread of scores. Identification scores were above 80% for only the M1 and the M2; the score for the BMP was the lowest (53%). The discrepancy between identification scores at the near and far ranges varied with vehicle. There was little difference between the two ranges for the M1 and the M2; a much larger difference in favor of transferring from far to near images for the other vehicles. The test format also influenced the scores, with the Vehicle ID test yielding slightly higher scores than the Signature Challenge format.

- Does thermal imagery training transfer to visible imagery regardless of the training condition?

- Does thermal imagery training transfer to vehicles not trained regardless of the training condition?

Experiment 3 replicated the pretest-posttest effects in the other two experiments. The scores for vehicles that received the FLIR training increased from the pretest to the posttest. Scores remained the same for vehicles that were not trained. This pattern occurred for both

visible and FLIR imagery. Enhanced skill and confidence with the FLIR imagery were also demonstrated by the response times, as they were cut by more than half on the posttest.

The Starting Point

The soldiers in this experiment could be considered representative of soldiers in active component Mechanized Infantry and Armor units. In general, they had experience with thermal sights and considerable field experience, unlike the soldiers in the other experiments. However, like the soldiers in Experiments 1 and 2, they had little formal combat vehicle identification training.

Their experience from being in a Mechanized unit was reflected in the pretest scores. They scored well on US vehicles, both visible and FLIR images, common in such units: the M1, M2, HMMWV, and the M113. Their lack of training on and experience with nonArmy and foreign vehicles was also reflected in these pretest scores: the Marine Corps LAV, the T72, BMP, BTR, and ZSU. As in the other experiments, the soldiers were highly motivated throughout the training.

The Learning Process

Results on the engine and exhaust test indicated that more emphasis should have been given to the symmetrical and unsymmetrical locations of the engine and exhaust. Higher scores were expected on this test, given the extent to which the instructor stressed the criticality of these features to thermal vehicle identification. Some means of pointing to these features in a very clear and systematic manner in the program's software, by use of directional arrows or other graphic aids, may be needed. Also in this experiment, instructional time was not devoted to directly comparing vehicles with similar thermal signatures. Such instruction could have reduced the errors on engine and exhaust locations, particularly for vehicles with which soldiers were less familiar.

The scores on the first training module, Vehicle ID, reflected the confusions one might expect with limited vehicle recognition training. Clearly, these confusions were more evident at the far range band and with the less well-known vehicles. In addition, the difficulties soldiers had with the front and rear views throughout the experiment were apparent during this first training phase. Interestingly, problems in identification at the longer range band, with the more difficult vehicles, and with the more difficult front and rear aspects, were associated with slower times to respond. These findings are consistent with Briggs and Goldberg's (1995) research using line drawings of vehicles.

The time spent on the Vehicle ID training module indicated that soldiers took seriously our guidance to study the images and learn from the corrective visual feedback. The typical response time was 23 sec for the near images and 32 sec for the far images. An additional 4 to 5 sec elapsed before going to the next vehicle. Response times in Signature Challenge were typically under 2 sec, even though 10 sec was allowed. The immediate corrective, visual feedback was found only in the Vehicle ID training module. It appeared that training first on this

module reduced the number and length of the learning plateaus in Experiments 1 and 2. If a gated, sequential version of Vehicle ID had been available, the process of learning to identify the vehicles could probably have been enhanced.

The series of Signature Challenge training exercises demonstrated clearly that when vehicles were presented at the far range band they were more difficult to learn to identify than when presented at the near range band. The number of sessions to criterion for the far range band was twice that for the near range. This ratio is similar to that found by Haverland and Maxey (1978), which was 1.5 times as many trials for simulated ranges of 3000 and 4000m using 13x optical sights. However, the relative distance between the two ranges in the Haverland and Maxey study was less than here. In Haverland and Maxey the ratio of the longer to shorter distance was 1.3; here, the corresponding far to near ratio was 2.4. The CVI*Plus* program could be improved by including specific feedback on vehicle confusions (e.g., which aspects of which vehicles are creating problems), coupled with a remedial capability. Remediation would allow soldiers to retreat to the Image Library for self-study of the vehicles they are confusing, or to another instructional module. In addition, if the Vehicle ID module could have been gated with a second exercise containing the oblique aspects, the number of Signature Challenge sessions required for mastery at the far ranges might have been reduced, but probably not eliminated.

The Signature Challenge training results clearly reinforced a finding from Experiment 1. One cannot assume that training soldiers to discriminate Vehicle A from Vehicle B means that they can also discriminate Vehicle A from Vehicle C. This problem was shown in the last two exercises in Signature Challenge. Soldiers quickly met criterion when all the US vehicles were suddenly grouped together, but had difficulties with a parallel grouping of nonUS vehicles. This occurred despite soldiers achieving an identification standard of 85% correct for these vehicles when presented in US and nonUS pairs. One explanation for this finding is that the nonUS vehicles are similar in many respects -- low profile, many parts appearing relatively cool, similar fronts and rears. On the other hand, the US vehicles are more distinct thermally. Smith et al. (1980) also found a decrement in performance when main battle tanks were unexpectedly combined in a special training session. In addition, the M1 and M2 were familiar to this group of soldiers, as indicated by high pretest visible scores.

A training strategy that deliberately presents different mixes of vehicles and vehicle imagery over a series of training exercises may be needed to ensure soldiers acquire a fairly robust concept/picture of vehicles. Soldiers must learn to discriminate vehicles from many possible "lookalikes" or "confusables." This finding also has implications for testing and the degree of skill assumed on the basis of tests. Clearly, vehicle identification scores can be affected by the pool of vehicles in the test. With a pool of lookalike vehicles, scores are likely to be low. With a pool of distinct vehicles, scores are likely to be high.

During training, soldiers also responded faster to vehicles they knew the best and to the easier aspects (flanks as compared to the front and rear). This relationship occurred on both the Vehicle ID and Signature Challenge training exercises.

The Outcome

The training increased vehicle identification skills. Scores on tests following training showed that soldiers were able to apply their skills to imagery and vehicle pools that differed from the training set.

Proficiency in identifying vehicles when shown the most difficult views, the front and rear, did increase as a result of the training. Skills with the near imagery were maintained upon testing these aspects within a larger imagery pool. However, skills with far imagery decreased somewhat from the maximum levels achieved during Signature Challenge training. This decrease could reflect the difficulty of this imagery, cognitive fatigue, confusions created by the imagery pool that included oblique aspects at the near range band, or a combination of these factors. The confusion matrices for the tests showed that the vehicle confusions involving the front and rear views had narrowed. They no longer reflected the generalized pattern of interference among vehicles and aspects evidenced during the first Vehicle ID training session. The confusions were confined to specific vehicles. At the test stage, most confusions occurred when the front of the BMP and the BTR and the rear of the LAV were presented. In addition, the confusions were replicated in both transfer tests, regardless of the different test formats.

Transfer of skill was tested with the oblique aspects. Soldiers did transfer skills from near to far imagery and vice versa. The easiest transfer was from the far to the near imagery. More difficult was the transfer from near to far imagery. The level of performance was at least 75% correct at the near range for all vehicles. However, for the far imagery, this level of proficiency was achieved on two vehicles only, the M1 and the M2. For the other vehicles, identification at the far range band varied from 53% to 68%.

Although the number and type of vehicles in the experiment were limited, the results indicate likely confusions as the population of vehicles increases. For example, the two tanks tended to be confused with each other. One identifying feature of the T72 is its exhaust, which is toward the rear on the left side. When soldiers viewed oblique aspects of the T72 that did not involve this side, they tended to confuse the T72 with the only other tank in the test pool, the M1, and in some cases, with the only other low profile nonUS tracked vehicle, the BMP. The BMP tended to be confused with the only other tracked APC in the test set, the M2, particularly at the long ranges. These two APCs have similar heat plants, as do many other APCs. The engine is to the right front and the exhaust is on the right side. This similarity probably contributed to confusions during training and upon transfer. The number of vehicles that look alike will increase as the vehicle pool increases. Instructional and training strategies that focus on addressing specific confusions such as these are needed.

The Vehicle ID and Signature Challenge tests also showed continuing difficulty with identifying vehicles other than the M1 and the M2, when vehicles were at distance and when the front and rear views were shown. Identification scores tended to be lower and response times slower. The Vehicle ID test results also showed that determining the aspect angle was problematic for the front and rear views and some front oblique aspects were misidentified as a flank, particularly at a distance.

Vehicle confusions were similar on both tests, Vehicle ID and Signature Challenge. However, identification scores were generally higher for Vehicle ID, probably because it was not timed, and soldiers could change their response before progressing to the next image. There was also a tendency for Vehicle ID scores to be higher when the imagery was problematic or challenging; with vehicles other than the M1 and the M2; and when soldiers had the same ranges in instruction and practice with no exposure to imagery at different ranges prior to testing.

Experiment 3 replicated the pre-post effects in the other two experiments. The scores for vehicles that received the FLIR training increased from the pretest to the posttest; scores for vehicles that were not trained remained the same. This pattern occurred for both visible and FLIR imagery. It is important to note that the pretests and posttests presented 10 vehicles, but there were response buttons for 14 vehicles. If the buttons had been limited to 10, then the scores for the nontrained vehicles might have increased because soldiers would be able to narrow their options considerably. The response format was probably a better indication of skill and knowledge, as real-world combat conditions are not multiple-choice. Retaining some ambiguity in the choice of vehicles in these tests was a more stringent assessment of skill.

Enhanced skill and confidence with the FLIR imagery were also demonstrated by the response times. Times were cut by more than half on the posttest.

The initial expertise of the soldiers was typically associated with posttest performance, despite the overall increase in vehicle identification proficiency. When soldiers were tested on the images used in training, the best predictors of performance were indices of training (training scores and rate of progress). Yet when tested for transfer to a different range, the soldier's initial expertise was the best predictor of performance.

The increased time and trials required to train the far imagery to the same standard as the near imagery raise several issues. First, although it is desirable to have soldiers look for cues that can be discriminated at the approximate range of their weapon system, it may not be desirable to train soldiers to identify vehicles initially at these longer ranges. The training took twice as long and transfer did occur, although it was better for some vehicles than others. Second, additional instructional techniques might enhance transfer from near to far. The instruction could focus on how thermal signatures change with range. It could also include guidance on how to change the brightness and contrast settings to bring out discriminating cues at range. Collectively, these distinct approaches would provide the soldier the information needed to determine the vehicle. Third, the effectiveness of different part-task training strategies should be investigated. Exercises could progress systematically from the nearest to the farthest range, or farther ranges could be combined gradually in the vehicle sets and shorter distances eliminated, as soldiers progressed from near to far images. Another option would be to train on the intermediate ranges. This might be more efficient when training time is limited, than training on only the far ranges as was done in this experiment.

SUMMARY

Summary and Discussion of Findings

The Value of FLIR Training

- FLIR training improved soldier expertise in identifying the vehicles in the training experiments. This expertise did not transfer to other, non-trained vehicles.

FLIR scores increased substantially from pretest to posttest in each experiment, an average of 34%. In addition, scores on the visible images of the same vehicles also increased in each experiment, an average of 22%, indicating that training on FLIR had the secondary benefit of improving skills with visible images. In contrast, FLIR and visible scores for the vehicles not in the training program increased, on average, only 6% and 3%, respectively. Greater skill was also shown in faster posttest response times to FLIR imagery. Posttest times were typically within 2.5 sec of the reaction times to the visible images, compared to a difference of 6.72 sec on the pretest.

These findings are particularly significant as they were replicated in different settings; with different samples of soldiers, with different training strategies, and with variations in the vehicle sets. Soldier expertise with vehicles (visible and thermal imagery) ranged from those with minimal skill to those with considerable expertise. The training strategies varied on several dimensions: exercise format (timed, not timed, type of feedback), simulated vehicle range, training criterion, size and sequence of vehicle sets. Either six or eight vehicles were trained. Five vehicles were common to all experiments: the M1, T72, M2, BTR, and LAV. In addition, the ranges at which the vehicles were trained did not always correspond to the close-up range used on the pretests and posttests. Clearly, there was no intent to train on only the images in the final test.

Given the variation in training conditions, it is not possible to pinpoint one factor that accounts for these results. Several possibilities should be mentioned. One, the instructor was thorough and systematic in the presentation of thermal cues; stressing heat signatures visible from most aspects and ranges, and pointing out how to discriminate vehicles with similar signatures. Use of the two-display and three-display screens in the CVI*Plus* Image Library facilitated the instructor's ability to relate visible and thermal images of the same vehicle and to compare the thermal signatures of different vehicles. Two, in all exercises the soldiers were active, not passive, and they progressed at their own rate. Three, training to a given criterion before progressing to the next vehicle set was important. Criterion attainment may have reduced the likelihood of interference that typically occurs when new vehicles are prematurely added to the training. Four, without the trial-by-trial feedback, learning would have been much less effective. Fifth, the soldiers were motivated. Invariably they indicated the need for this training, and that they had no other training materials that met this battlefield requirement.

Impact of Soldier Expertise

• Initial individual differences persisted throughout training and testing, despite substantial improvement in thermal skills for everyone. Soldiers with higher initial skills tended to score higher on initial sessions of the practical exercises and on tests. They also progressed at a faster rate through the training.

Correlations among pretest and posttest FLIR and visible scores were high and significant. Over the three experiments, of the 18 correlation coefficients between these scores, 17% were greater than .90, 61% were greater than .80, and 89% were greater than .70. In addition, pretest scores tended to correlate with progress on the training sessions; the higher the soldier's pretest score, the fewer sessions required to achieve the training criterion on the vehicle sets. In Experiment 2, initial skill levels also correlated with both the transfer and no transfer test scores. However, in Experiment 3, initial skill correlated with proficiency on the transfer imagery, whereas indices of training progress correlated with test scores on the no transfer imagery.

These findings indicate that individual differences in most test scores did not change substantially as a result of training. Even though all soldiers became more skilled, those at the top tended to stay at the top. There was still room for improvement on the posttest and on transfer tests. Not everyone "maxed" the tests, although some soldiers did score 100% on several tests or missed only one or two vehicle presentations.

Aspect Angles

• Some aspects were easier than others to determine. The aspects that were easiest to identify typically made it easier to identify the vehicle as well. Furthermore, training on vehicle aspect improved aspect scores.

The flanks were easiest to determine, followed by the oblique aspects, with front and rear views the hardest, particularly at the longer ranges. A corollary finding was that higher identification scores were associated with the easier-to-determine aspects. In other words, when the vehicle's flanks were presented, vehicle identification scores were higher than when the front and rear were displayed. This difference was typically larger for vehicles the soldiers found particularly difficult to identify overall. In addition, reaction times were faster when soldiers knew the vehicle, its aspect, or both dimensions. These findings both support and expand those of other researchers. None of the prior research on vehicle recognition and identification used eight aspect angles. Nor was reaction time recorded in the previous research. Thus our findings provide a more complete understanding of the mediating role of aspect on determining a vehicle's identity and the speed with which this is accomplished.

It must also be noted that there will be exceptions to these general findings regarding aspect. Some vehicles have distinctive rear thermal signatures. Some have distinctive signatures from the rear oblique or front oblique, making it easy to determine the orientation of the vehicle. For other vehicles this is not the case.

In addition, the aspect results do not simply reflect the amount of the vehicle's exposed surface area. If so, then oblique aspects would be easier than flanks. In reality, it is often difficult to determine the vehicle's orientation with an oblique view. Obtaining a mental image of how the thermal signature of the exhaust on the vehicle's flank changes with rotation is not easy. The shape of the hot spot from the exhaust when viewed from an angle varies with the size, shape, and exact location of the exhaust itself.

Lastly, our results showed that training on aspect was helpful, though determining aspect was easier than identifying vehicles. In Experiment 2, we found that encouraging soldiers to think in terms of vehicle orientation had beneficial results. Those who practiced determining the vehicle's aspect, as well as its identity, were more likely to determine aspect correctly on transfer imagery than soldiers who did not have aspect training.

Training Large Numbers of Vehicles

- Vehicle sets with many images led to learning plateaus and learner frustration. Shorter and multiple sets worked better.

For each vehicle, views of eight aspect angles were available. In prior vehicle identification research using photopic displays (Warnick & Smith, 1989), the maximum number of aspects used was five (no rear and rear oblique views). During training we used all eight aspects in order for soldiers to acquire a three-dimensional concept of the thermal signature of each vehicle. This number of images directly impacted the length of the training exercises. In Experiment 1, sessions with many images (e.g., 8 vehicles with 64 images) demanded a high level of concentration. This situation was frustrating when the soldier did not reach criterion and had to go through the entire set again. If summary feedback by vehicle and a remedial training option had been available, fewer learning plateaus might have occurred.

We developed a rule of thumb for the other experiments that no more than 48 images, preferably no more than 36, would constitute a vehicle set. In Experiments 2 and 3, we reduced the number of vehicles from eight to six, and incorporated multiple, short vehicle sets whenever possible. This procedure worked well.

Performance Standards

- There are training costs associated with training to a high standard initially or training with difficult imagery. For some soldiers, very high standards, above 90%, cannot be achieved within a short training period, particularly for difficult imagery (e.g., at far ranges, for similar vehicle sets, difficult aspects). Learning plateaus are common in such situations. Although soldiers did learn to discriminate vehicles at far ranges, starting training on distant vehicles was not an efficient strategy for those with limited initial skills, even though transfer to near ranges was high.

The primary cost associated with high training standards or difficult imagery is that of time. Some of the difficulties in reaching criterion in Experiment 2 may have been associated

with the fact that a long range was used for training. The results of Experiment 3 showed clearly the impact of long distances on training efficiency. Soldiers required twice as many training sessions and twice as much time to complete the training exercises on the far as opposed to near imagery. Soldiers trained on the far imagery transferred their skills well to near imagery. In contrast, soldiers trained on the near imagery had trouble when transferring to far imagery. Thus an instructor must make a trade-off between the longer time required to master far ranges and the resultant high ability to identify vehicles at closer ranges, and the shorter time required to master near ranges concomitant with less ability to identify at far ranges.

- Results showed clearly how the image pool, the training criterion, and the response format can affect performance on tests and learning rate. These findings have implications for testing, and the degree of skill assumed from training and test scores.

Obviously, test scores can be affected by the pool of vehicles. With a pool of similar vehicles, scores are likely to be low. With a pool of distinct vehicles or very familiar vehicles, scores are likely to be high. Vehicles displayed at long ranges will be harder to identify than those at closer ranges. New imagery will be more difficult than the training imagery. Training progress and test scores are also influenced by the size of the image pool. For example, the probability of a correct response is greater for three vehicles than for ten as the pool of potentially confusing vehicles is smaller. On the posttests, our findings would have been quite different if vehicles that were not trained had been excluded. Both test standards and training criteria need to consider these factors.

Similarly, several factors influence the ease of progressing from one vehicle set to the next. Lowering the training criterion could mean that soldiers are progressing to other vehicles without first mastering the current vehicle set. In establishing a training standard, the number of images must be considered. Is the standard established so that a soldier can miss many images of one vehicle in the set and still pass? For example, three vehicles with eight aspects per vehicle constitutes an exercise of 24 images. If the training criterion is 80%, a soldier could miss five of the eight images of one vehicle, yet still pass by getting all images of the other two vehicles correct. In our experiments, the training criterion was initially established on the basis of an absolute error rate, rather than a fixed percentage. Therefore the percentages varied with the number of images. We did not want soldiers to progress to the next vehicle set if they could miss many images of a vehicle. On the other hand, we found that when the criterion was very high (above 90%) on a large number of images (48 to 64), soldiers became discouraged on repeated attempts to reach this criterion.

Time limits established for responding must consider the time required for the soldier to interact with the computer display and the computer equipment. These factors are in addition to the time needed to identify the vehicle itself. Longer times could be set for training than for testing. Over the course of the experiments we found that ten seconds was appropriate for displays of three and four vehicles in Signature Challenge. Ten seconds for eight vehicles and five seconds for two vehicles were too short. Soldiers had insufficient time to study the images when they also had to locate the correct response button.

Ultimately, the training criteria and test standards are the instructor's or the training organization's decision. Instructors need to be intimately aware of these trade-offs when establishing standards, rather than automatically applying a commonly used criterion. To have a fair, yet rigorous assessment of proficiency, decisions on standards should be made in the context of the difficulty of the imagery and other factors that have been shown to influence soldier performance on thermal vehicle identification.

Handling Vehicle Similarities and the Resulting Confusions

- Even after extensive training to criterion, some vehicle confusions remained, for certain aspects and at certain ranges. To ensure expertise in vehicle identification, training strategies must consider the similarities in thermal signatures regardless of the vehicle class (e.g., tank, APC, logistics). And instructional techniques must focus on the vehicles that are hard to discriminate.

Findings from each experiment supported this conclusions. Consider the following examples of confusions that occurred after training to a high criterion. First, in Experiments 1 and 2, soldiers trained on part-task vehicle sets and completed training with a set of all vehicles. In both situations, when soldiers were exposed to all the vehicles, cross-set interference occurred. Second, in Experiment 3, soldiers were first trained on US and nonUS vehicle pairs. Then the vehicles were regrouped into a vehicle set composed of all the nonUS vehicles and one with all the US vehicles. While soldiers easily discriminated the US vehicles, this was not the case for the nonUS vehicles. The number of sessions required to distinguish the nonUS vehicles increased, particularly at the far ranges. It was hypothesized that the similar thermal signatures of the nonUS vehicles created confusions, whereas the thermal signatures of the US vehicles were relatively distinct. Third, in Experiment 3, the transfer test included both transfer and no transfer images of the vehicles. Despite repeating some of the training imagery in the test, proficiency with this imagery was not always maintained when new, not previously trained images were included.

The findings support Lintern's (1989) statement that high levels of skill on part-tasks do not always mean that individuals have developed resistance to interference from additional loads. In the first two examples cited above, we found that you cannot assume if soldiers learn to discriminate vehicle A from B and vehicle D from E, that they can automatically discriminate A and B from D and E. Nor will their level of expertise in discriminating the original vehicles be maintained when other vehicles are added to the pool. Soldiers must learn to discriminate vehicles from their "lookalikes." An instructional and training strategy is needed that deliberately presents mixes of similar vehicles and also considers the most likely confusable vehicle image conditions (e.g., aspects, far distances, day).

These confusions might have been reduced if an instructional module had been available in the CVI*Plus* prototype that systematically pointed out thermal features of each vehicle. One approach to this instructional module would be to juxtapose the visual and thermal signatures, highlighting key thermal features by the use of arrows, audio instruction/elaboration, or both, as the vehicle rotated 360°. This technique would pinpoint the principal hot and cool spots while simultaneously linking them to the heat plant of the vehicle and to other physical structures.

Just as important, however, is that persistent confusions remaining after the "standard" training point to the need for special instruction on what might be termed "the hard ones." In most cases, it was relatively easy to determine, after the fact and based on a close examination of thermal signatures, why these confusions remained with the imagery used in our experiments. However, it is questionable whether these confusions could have been predicted beforehand. An advanced training program would, nonetheless, include an optional instructional module on confusions, to be selected by the individual at any time during training. This special instruction would systematically address the vehicles and the associated conditions that make discrimination difficult. Side-by-side comparisons of "confusable" images would be pivotal to the effectiveness of this module. This instruction would not focus on vehicles in isolation nor on vehicle conditions that are easy to discriminate.

Facilitating Transfer

- Not all transfer conditions were equal. It was easier to transfer from far to near images than from near to far. Transfer to black-hot night imagery from white-hot night imagery was easier than transfer to black-hot imagery during the day. Transfer was easiest when flanks were presented, followed by the oblique aspects, and hardest with the front and rear views.

The degree of transfer can be explained primarily by theories of transfer. The more identical or similar the elements, the easier the transfer. The more changes in the similarity of the images, the harder the transfer. To the degree that information encoded during learning was retrievable with cues available in the transfer situation, then positive transfer occurred. If the encoding of the training material was incompatible in some way with information or material in the transfer task, then less transfer occurred.

To illustrate these concepts, consider training on white-hot night imagery. When soldiers had to transfer to black-hot night imagery, basically one dimension changed, that is, the "color" of the hot spots. One image could be described rather generically as a "negative" photo of the other. On the other hand, when going from white-hot night to black-hot day, not only did the color of the hot spots change, but many of the hot spots themselves changed. Some hulls and turrets that appeared cool and almost invisible at night appeared hot and very distinct during the day. Suspensions that were very hot at night relative to the rest of the vehicle were cool in the day thermal images. In addition, no instruction was provided on the thermal cues that appear during the day. Soldiers were trained to attend to the "night cues." Consequently, what was learned on the original task of white-hot night imagery was, in some instances, inconsistent with the black-hot day imagery on the transfer test.

- The best exercise format for training a specific set of images may not be the best for facilitating transfer or for initial learning. Techniques such as knowledge of performance with visual corrective feedback and self-paced, nontimed exercises were beneficial. They seemed to help soldiers understand their confusions, made learning more efficient, and enhanced transfer. Drill-type exercises may be more appropriate for "fine tuning" skills, than for initial learning and transfer.

The ability of soldiers to transfer vehicle identification skills to new imagery and new situations is critical to success in combat. It could be argued that high levels of transfer should be the major goal of training programs, rather than mastery of a specific set of vehicle images. Findings from Experiment 2 showed that transfer was enhanced by a training format that provided them with visual, corrective feedback, and forced them to study and examine the thermal images. Using such a strategy, soldiers did not simply memorize the imagery, but came to understand their errors at a deeper level. A training format that demanded quick responses and provided the typical knowledge of results was not as beneficial, as it afforded soldiers only a cursory understanding of their mistakes and confusions.

These findings are consistent with much of the literature on transfer. In particular, it is consistent with those who postulate a trade-off between training for transfer and training for rapid acquisition of the training tasks (Cormier, 1987; Cormier & Hagman, 1987; Salomon & Perkins, 1989). The "low" road to transfer (Salomon & Perkins, 1989) is conceptualized as one based on extensive practice and automatization on the training task. The "high" road is conceptualized as one based on mindful abstraction and a deeper understanding of the training task, which results in slower attainment of transfer but the transfer is more extensive and adaptive in nature.

Our experiments did not allow a direct test of the relative merits of a training approach that overtly stressed understanding the heat plant of each vehicle, and how the vehicle's thermal signature generated by the heat plant changes with aspect and range. However, if transfer is the training goal, instruction with close-up imagery is key to soldiers understanding the heat plant of vehicles. For example, it is only at close ranges that one can distinguish the shape of the external, muffler-like exhaust system on the right flank of the LAV and how the exhaust signature carries over to the front and to the right oblique views. It is only at close ranges that one can see the differences in the thermal signatures of the rear of the LAV and the M2, and consequently appreciate how these differences are viewed at range. Only at close ranges can the separate exhaust grates on the rear of the M60 be seen, and how the thermal signature in the rear appears different from other tanks with rear engines. Lastly, side-by-side displays of the thermal and visible images greatly assist in this instructional process, particularly for vehicles soldiers have not seen before.

In Experiment 2, the non-timed training exercises that provided side-by-side corrective visual feedback (the Vehicle ID module) tended to result in a steady progression toward the training criterion for soldiers who had difficulty with the imagery. On the other hand, the fixed-pace, timed exercises where only "correct" or "incorrect" feedback was provided did not seem to have these benefits, although the training criterion was eventually achieved. The two training formats could have sent different cues to the soldiers about the intended outcome of the exercise: to respond correctly and fast versus to respond correctly based on a careful examination of the imagery. The drill-type exercises may have been perceived as being more like a "contest" than a learning environment. This interpretation is consistent with Salomon and Globerson's (1987) argument that training and instructional materials can influence the learner's approach to a task.

Value of Computer-Based Training

- The computer-based training format was efficient. The training modules adapted to individual differences in initial skill and in rates of learning.

Generally speaking, a computer-based program, which allows individuals to progress at their own rate and tailors feedback to the idiosyncratic responses of each individual, can be more efficient than classroom instruction, which does not adapt easily to individual differences. In our experiments, the total training and testing time per vehicle typically averaged 20 minutes, ranging from 12 to 27 minutes for the fastest and slowest individuals respectively. It should be noted that these times also reflect the group instruction time on vehicle cues.

Previous research (Gibson, 1947; Warnick & Smith, 1989; Whitmore et al., 1968) using group instruction techniques required more time to train identification skills. For example, Whitmore et al. found that it took 16 hours to train all Air Defense Artillery trainees on 16 aircraft to the desired criterion of 95% (60 minutes per aircraft). The fastest trainees required only 11 hours. During World War II, Gibson and colleagues reported various lengths of classroom training: 26 to 30 hours on 40 planes (40 to 45 minutes per aircraft). In general, in the combat vehicle identification studies at Ft. Hood, TX (Warnick & Smith, 1989; Smith et al., 1989), 150 minutes was used for a set of 5 vehicles (30 minutes per vehicle).

Although faster, it would be unwarranted to conclude that computer-based instruction is twice or three times as fast as group instruction. There was insufficient information to make such quantitative comparisons with the prior research. Nonetheless, the results point to the benefits of computer-based technology for training these skills. Even greater efficiencies could be achieved with enhancements that optimize the training modules and training sequence.

Soldier Reactions

- Soldier reactions to the program were positive.

Based on comments made by the soldiers (see Appendix E) and observations during the experiments, the program was a success and was viewed as meeting a critical training need. The self-pacing features, individually tailored feedback, and the automated gating procedure with increasingly difficult exercises or different sets of vehicles were intrinsically motivating. These techniques also made soldiers appreciate the challenge of vehicle identification. Use of actual thermal imagery in conjunction with multi-media, computer-based, practical exercises that adapt to each soldier's learning rate and can continually challenge soldier expertise is clearly the way to make the difficult task of vehicle discrimination more interesting, efficient, and effective.

Program Features as of July 1997

Since our experiments, changes have been made to the prototype CVI*Plus* program. These changes reflect the findings from our experiments, soldier feedback from beta sites, and other research efforts conducted by the NVESD. In July 1997, the program described below was formally delivered to the sponsor, PM-FLIR. However, it is under continuing development,

expansion, and refinement as additional resources become available. The database has expanded to 36 vehicles. These vehicles, by type and country of origin/use, are in Table 4.

Table 4

Vehicles in the CVIPlus Training Program as of July 1997

Type of Vehicle	US	NonUS
Tank	M1A1 Abrams M60A3 M551 Sheridan	T72 (CIS[a]) T72 with reactive armor (CIS) T62 (CIS) T55 (CIS) Challenger 1 (UK) Leopard II (Germany)
IFVs, Tracked APCs, and Reconnaissance Vehicles	M2/M3 BFV M113	BMP-1 (CIS) BMP-2 (CIS) MT-LB (CIS) Marder 1 (Germany) Warrior (UK)
Wheeled APCs and Reconnaissance Vehicles	LAV-25 LAV-Mortar LAV-AT M93 Fox (also Germany)	BTR-70 (CIS) BTR-80 (CIS) BRDM-2 (CIS) Spahpanzer Luchs (Germany)
Field Artillery (self-propelled howitzers)	M109 155mm Palladin	2S1 123mm (CIS) 2S3 152 mm (CIS) AS90 155 (UK)
Air Defense Artillery		ZSU 23-4
Logistics	M998 HMMWV FMTV Medium Cargo Vehicle HEMMT M35 2.5-ton truck M814 5-ton truck	ZIL 131 truck GAZ-66 truck
Total	15	21

[a] CIS stands for Commonwealth of Independent States; used by *Jane's* to refer to Warsaw Pact vehicles.

The FLIR pretest and posttest now allow the soldier to change polarity to help identify the vehicle. This was done because most thermal sights have reverse polarity and soldiers use this capability in the target acquisition process. The posttest can only be taken when the soldier completes training on the vehicles designed by the instructor. This could be all vehicles in the program or a portion of them, depending on the instructional intent.

The instructional module entitled "Thermal Basics" has changed. This module now covers basic concepts regarding what FLIR energy is and what FLIR sights do (they see heat in vehicles, scenes, people). The terms thermal signature and cues are defined, and videos showing how signatures of a tank and an urban scene vary over a 24-hr day-night cycle are presented. The video "morph," formerly in the ID Training module, is now part of Thermal Basics.

Exercises on adjusting brightness and contrast, and on searching for targets and then attempting to identify them have been added. Short videos of 13 moving vehicles as seen through a thermal sight are available. The effects of obscurants on a thermal image are depicted. Vehicles in defilade positions are shown. Two exercises on vehicle detection or search are included, where the soldier gets practice in using the brightness and contrast controls to find vehicles of military interest, and can then attempt to identify them.

The summary feedback on the tests as well as the training exercises has expanded. Total percent correct is still provided. However, a vehicle by aspect matrix is displayed. The rows of the matrix are the vehicles; the columns are the eight aspects. Correct answers are indicated by green cells; incorrect answers by red cells with the name of the vehicle which was confused with the displayed vehicle. So if a soldier incorrectly identifies the LAV as an M2 when the rear of the LAV is displayed, the rear aspect cell for the LAV will be red with the abbreviation "M2" inserted. For the training exercises, the soldier can also "click" on each red cell and go to a display that compares the two vehicles that were confused.

The focus of Vehicle Basics has changed from a module on vehicle names to a module that presents information on vehicle cues and contains practice exercises on identifying vehicles. Vehicle Basics is now the primary instructional module within the program. Vehicles are presented in vehicle sets. Within each set, audio is used to describe the thermal cues for each vehicle. Both thermal and visible images of the vehicle are displayed on the screen. The aspect displayed corresponds to the audio description of cues, although the individual can manually control the aspect via the eight aspect buttons at the bottom of the screen. After the cues for a vehicle are presented, soldiers complete an aspect matching exercise, once for night imagery and once for day imagery. After two vehicles are presented, the soldier must execute either a Signature Challenge or Vehicle ID gated exercise to start training on discriminating vehicles within the set. This process continues through all vehicles in the set. In addition, the soldier has access to the Image Library for those vehicles on which he has been trained.

The Image Library remains. The Signature Challenge, Vehicle ID, Aspect Matching, and Vehicle Matching exercises also remain. Vehicle ID exercises are now gated. The Vehicle Comparison module has been removed. Sensor Controls is a separate module, rather than being embedded in Thermal Basics. The M1A2 Abrams sight is presented with information on and demonstrations of what happens with the brightness, contrast, polarity, magnification, and focus controls.

An Instructor Control Matrix (ICM) has been added to allow the instructor to select the vehicles and vehicle sets to be trained and tested. The instructor can determine how difficult to make the test. The maximum number of vehicles in a set is six. In addition, the training criterion and the exposure time (in Signature Challenge) are established through the ICM. Test results are summarized in this module as well.

The soldier has a choice of a fixed route through the training package or a self-determined route. If the fixed route is selected, he must go through program modules in the following order: Pretest (both FLIR and visible), Thermal Basics, Vehicle Basics, and Posttest (both FLIR and visible). If this route is not selected, then the order in which these modules are

used, as well as when the practical exercises and sensor control modules are selected, is controlled by the user.

These changes incorporate many of the primary suggestions for the program as determined through our research. They will enhance the program's effectiveness and improve its flexibility. A primary lesson learned from the experiments was the need for flexibility in the software to meet the spectrum of individual and instructor training requirements in both the field and the institution. The changes include: (a) creation of vehicle sets with limited numbers of images, (b) vehicle sets that can be programmed in terms of difficulty, (c) feedback summaries for Signature Challenge and Vehicle ID are tailored to each vehicle, (d) the option of varying the pass criteria to correspond to the soldier's expertise and training progress, (e) making each of the primary training modules gated, (f) an instructional module on vehicle cues, and (g) allowing soldiers to exit exercises to go to other modules such as the Image Library for additional study. An instructional block that focuses on the most difficult discriminations has not been included at this time.

Questions for Future Research

Additional research questions arising from the experiments are cited here. Obviously, the three experiments addressed a finite number of questions. Furthermore, some areas of interest could not be addressed because of the limited number of images available and other restrictions in the prototype software. Many decisions about the structure and content of the training program have, of necessity, been made without the benefit of empirical validation. However as we found in our experiments, one cannot always predict the outcomes of training techniques and procedures. Therefore, this final section addresses questions we feel should be investigated. The questions have both pragmatic and conceptual implications.

Part-task Training

- What are the effects of different part-task training schedules, specifically pure-part, repetitive-part, and progressive part?

We did not compare different part-task training schedules. We examined only variations of a pure-part schedule (Proctor & Dutta, 1995). Historically, other researchers (Gibson, 1947; Warnick & Smith, 1989; Whitmore et al., 1968) also used variations of the pure-part schedule. But the flexibility provided by computer software makes comparisons of different part-task schedules very feasible. Answers to the following questions would greatly assist training developers in designing future and similar programs.

Specific questions to address include which strategy leads to less confusion on the whole-task: pure-part, repetitive-part, or progressive-part? Which is the more efficient strategy? How effective is a pure-part schedule combined with remedial training tailored to the soldier's errors? Do repetitive and progressive part strategies lead to boredom? Are you likely to over-train on some vehicles and not get sufficient training other vehicles with repetitive- and progressive-part strategies?

- What cues or identifying features should be presented and how should they be presented? Does an emphasis on verbal descriptions of features or parts inhibit or facilitate making the appropriate perceptual discriminations? How does understanding the heat plant of vehicles facilitate the soldier's ability to discriminate thermal signatures?

These questions have not been systematically examined. The "what" and "how" questions are related. Researchers (Biederman, 1987; Gibson, 1947; Leibowitz as cited in Cockrell, 1978) agree that vehicle and aircraft identification training should focus on distinguishing features as opposed to descriptive features. If the distinguishing features of a thermal signature can be described adequately in verbal terms, then instructional techniques that are primarily verbal should be effective. However, if verbal descriptions fail to adequately depict the shapes characteristic of thermal signatures, then instructional techniques that stress perceptual learning are required. It may be that a combination of verbal and perceptual training strategies is needed, and that determining the balance between the two approaches over an entire course of instruction should be a focal point of future research.

As Gibson (1969) pointed out, a single property rarely determines the uniqueness of an object. Rather it is a set of characteristics that permits final identification, and often this set is difficult to describe. Combat vehicle identification training in the military typically relies rather heavily on verbal descriptions. This may stem in part from the limited number of vehicle images (pictures, drawings) typically available as training aids. Leibowitz (as cited in Cockrell, 1978, p. 3) argued against a heavy reliance on such training, maintaining that the ability to verbalize features comes after discrimination, not before, and that typically verbalization covers only part of the basis for the discrimination. He felt that military training should exploit the use of images, presenting many views, starting with easy discriminations and progressing to more difficult images. Cockrell (1978) did, in fact, find that reliance on verbal cues for target identification was not a particularly effective method of training image interpreters.

The inadequacy of the English language to describe objects completely and satisfactorily applies directly to the thermal signatures of vehicles. Words "work" part of the time, but not all of the time. They work for parts of an image, but not for all of it. With thermal imagery, it is relatively easy to describe the hot spots created by the vehicle's exhaust and engine and to point directly to them (e.g., the hot exhaust on the left flank of the T72 and T62; the stovepipe-like signature of the exhaust on the 2.5 and 5 ton trucks, the circular heat signature on the rear of the M1 from its turbine engine). But it is not easy to describe how these hot spots are configured in a three-dimensional space. Nor is it easy to describe the temperature differences generated by the rest of the vehicle that are typically less distinct, yet integral to its thermal signature.

What distinguishing features beyond the hottest spots in a vehicle's thermal signature, the engine and the exhaust, should be described? Would a framework for thermal cues such as S^4HEET[1] assist in determining these additional features? Can these features be described well verbally? When words are inadequate, what instructional techniques should be used? What

[1] Shape, size, suspension, symmetry of thermal signature, hull, engine, exhaust, and turret.

ranges are best for explaining and showing cues? Is special instruction needed to train day thermal signatures? Is special instruction needed for thermal cues at long ranges? When does attention to the vehicle's overall shape become relevant in the instructional process?

Is it best to train individuals on the distinguishing features of types of vehicles first (tanks, tracked APCs, wheeled APCs, howitzers), what has been called the basic level of abstraction (see O'Kane et al., 1997), and then focus on features that separate vehicles within categories? Expert military instructors often take this approach, because they know that soldiers must discriminate quickly among the types of vehicles on the battlefield when making target engagement decisions.

What are the benefits from describing how the thermal heat plant and material composition of vehicles influence the vehicle's thermal signature? Is learning more efficient, transfer better, or retention higher? When looking at a vehicle, individuals ask such questions as what causes this hot spot, what makes this part of the turret cooler than the rest of the turret, what makes this entire area appear warm, and why do the skirts appear cool. Most likely the answers to these questions come from a knowledge of the structural characteristics of the vehicle and its heat plant. For example, the hot spot may be the exhaust; the bustle rack on the turret may be causing the cool appearance, the warm area may be coming from the engine which is located behind other vehicle components; the skirts may appear cool because they are made from a rubber or vinyl-based material which hides heat at night. It would seem that less confusion among vehicles would result when soldiers have this knowledge.

What are the effects of different ways of presenting thermal cues in a multi-media system -- use of graphics, audio, and images? How effective are point and describe/label techniques? What techniques provide the soldier the best three-dimensional understanding of the vehicle's thermal signature? What is the best balance between presentation of cues and presentation of images without cues over the course of instruction as soldiers gain expertise? What techniques do soldiers prefer, both initially and at later stages of training?

Vehicle Sets

- Within the pool of vehicles of interest (as defined by the instructor or mission conditions), how should vehicle sets be created and how should they be sequenced? When should training on similar vehicles occur and how can similarity be operationally defined?

Our data showed it was important to have soldiers discriminate vehicles that present similar thermal signatures. However, in order to design an instructional program around similar sets of vehicles, one must first know which vehicles are, in fact, similar. What are the vehicle clusters? What are the factors that discriminate such clusters or sets? Are vehicle sets defined by traditional military vehicle categories or do they divide according to other dimensions? Do vehicles group the same regardless of whether the imagery is visible, thermal night, or thermal day? Do all aspects of a vehicle fall in the same cluster? What do vehicle clusters indicate about which aspects provide the best understanding of thermal signatures, and therefore which aspects should be used when training time is limited or limited views are available for training?

In order to structure a program around vehicle sets, there must be a method or analytic procedure that generates such sets. Although expert judgement has been used in the past to group vehicles or aircraft within training programs, are there analytic techniques that would automate the creation of such a space and also provide a means of easily determining how the vehicle clusters or sets change when new vehicles are added? Does such a method require empirical data on vehicle confusions? Can a valid method be developed based on subject matter experts' judgements, such as the decision-trees used by O'Kane et al. (1997) and Cockrell (1978)?

Our findings also raised questions about the sequencing of vehicle sets in training. At some point soldiers must discriminate vehicles that look alike. Even when soldiers were trained to a high criterion on pairs of vehicles, cross-set interference occurred when all vehicles were pooled. Very similar vehicles create difficult sets, and the greater the number of similar vehicles within a set, the greater the difficulty. On the other hand, if quite distinct vehicles are presented in the initial training, will soldiers process the imagery primarily at a superficial level, resulting in substantial confusions when vehicles are pooled? What are good compromise training sequences between these two extremes of very hard and very easy sets? Is special instruction needed that provides side-by-side visual comparisons of vehicles with similar signatures? When should this type of instruction occur (e.g., during the initial instruction, only for soldiers who demonstrate difficulty in discriminating similar vehicles)?

Adapting to Soldier Expertise

- Would soldiers with low initial skills benefit from a different training strategy than those with higher initial skills?

Specifically, would a self-paced format with visual, corrective feedback be more effective for soldiers with low initial skills than a fixed-paced knowledge of results format? Would a fixed-pace format with knowledge of results be better for soldiers with higher initial skills? Our findings indicated that self-paced, non-timed exercises with immediate visual corrective feedback that contrasted the confused vehicles enhanced transfer and helped soldiers who had difficulties. But is this format the best for soldiers with more expertise? Would their training time be better spent in sharpening their skills with hard confusion sets and timed exercises?

Answers to these questions would also help to address sustainment and refresher training issues. What is the best way to sustain skills? Should the initial training simply be repeated? Can diagnostic tests be developed that identify differences in expertise so training can then adapt to these differences?

Skill Retention

- How quickly do soldiers lose their vehicle identification skills and what factors affect the rate of forgetting?

The need to answer these questions should be obvious, as they impact what leaders assume about their soldiers' skills and how training schedules are developed. The question assumes soldiers have mastered vehicles to a specified criterion initially, as one can not forget something that was never learned in the first place. There are three major corollary questions of interest. Do soldiers tend to forget the vehicles they had trouble learning initially? Are these the vehicle confusions that reappear with time? And how do different training strategies affect the retention process? Little research has been directed at these issues (for photopic images see Heukeroth, Smith, & Shope, 1988), so our knowledge of the dynamics of the retention process is minimal.

Transfer of Skill

- To what extent are skills maintained as the differences between the training conditions and non-training conditions increase?

Transfer should be assessed in at least two different contexts: using imagery within the program, as was done in our experiments, and using tactical sights and vehicles in field exercises. With program imagery, the primary questions are whether the FLIR training provides soldiers with the skill to discriminate among the vehicles on which they were trained by using new or unfamiliar images of the same vehicles. The degree to which this new imagery differs from the training imagery can be varied (different ranges, different angles, defilade positions, combat-loaded vehicles, time of year, time of day) With program imagery, relatively gross discriminations can also be assessed. For example, can soldiers simply discriminate vehicles on which they were trained from vehicles never presented during training? However, the ultimate questions are whether the FLIR training transfers to field settings with actual vehicles and tactical sights, and if the training strategies that enhance transfer when using program imagery also enhance performance in field settings. These final questions are the critical ones, as the goal of the training is to prepare soldiers for combat.

REFERENCES

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115-147.

Briggs, R. W., & Goldberg, J. H. (1995). Battlefield recognition of armored vehicles. *Human Factors, 37*, 596-610.

Cockrell, J. R. (1978). *Evaluation of four target-identification training techniques* (Technical Paper 301). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A061 175)

Cockrell, J. T. (1979). *Effective training for target identification under degraded conditions* (Technical Paper 358). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Cormier, S. M. (1987). The structural processes underlying transfer of training. In S. M. Cormier, & J. D. Hagman. D. (Eds.), *Transfer of learning: Contemporary research and applications* (pp.152-181). New York: Academic Press.

Cormier, S. M., & Hagman, J. D. (1987). Introduction. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications* (pp. 1-8). New York: Academic Press.

Cullen, T., & Foss, C. F. (Eds.). (1996). *Jane's™ land-based air defence, 1996-97* (9th ed.). London: International Thomson Publishing.

Department of the Army. (1970, June). *Visual aircraft recognition* (GTA 44-2-1). Washington, DC: Author.

Department of the Army. (1977, February). *Armored vehicle recognition* (GTA 17-2-8). Washington, DC: Author.

Department of the Army. (1979, October). *Aircraft recognition* (GTA 44-2-10). Washington, DC: Author.

Department of the Army. (1984a, August). *Aviator's recognition manual* (FM 1-402). Washington, DC: Author.

Department of the Army. (1984b, October). *Armored vehicle recognition* (GTA 17-2-13). Washington, DC: Author.

Department of the Army. (1985, June). *Combat vehicle identification* (GTA 17-2-11). Washington, DC: Author.

Department of the Army. (1987a, May). *Armored vehicle recognition* (GTA 17-2-13, Change 1). Washington, DC: Author.

Department of the Army. (1987b, October). *Soldier's manual of common tasks: Skill level 1* (STP 21-1-SMCT). Washington, DC: Author.

Department of the Army. (1996). *Bradley Fighting Vehicle gunnery* (FM 23-1). Washington, DC. Author.

Donohue, J. (1991). *Introductory review of target discrimination criteria* (FL-TR-92-2129). Hanscom Air Force Base, MA: Phillips Laboratory, Air Force Systems Command. (DTIC No. AD-A254 954)

Foskett, R. J., Baldwin, R. J., & Kubala, A. L. (1978). *The detection ranges of features of armored vehicles* (Technical Report 78-A37). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A068-043)

Foss, C. F. (1992). *Jane's AFV recognition handbook.* Alexandria, VA: Jane's Information Group.

Foss, C. F. (Ed.). (1995). *Jane's™ armour and artillery, 1995-96* (16th ed.). London: International Thomson Publishing.

Foss, C. F., & Gander, T. J. (Eds.). (1996). *Jane's™ military vehicles and logistics, 1996-97* (17th ed.). London: International Thomson Publishing.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Gibson, J. J. (Ed). (1947). *Motion picture testing and research* (Report No. 7). Washington, DC: Army Air Forces, Aviation Psychology Program.

Hansen, O. K. (1970) *Airborne operator target recognition training procedures, Volume II. Visual reconnaissance study* (AFHRL-TR-69-34). Wright-Patterson Air Force Base, OH: Training Research Division, Air Force Human Resources Laboratory. (DTIC No. AD 374 810)

Haverland, E. M., & Maxey, J. L. (1978). *Problems in helicopter gunnery* (Technical Report 78-A356). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A046 274)

Heuckeroth, O. H., Smith, N. D., & Shope, G. L. (1988). *Target acquisition and analysis training system (TAATS): Retention and effects of retraining* (Research Note 33-103). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A204 294)

Heuckeroth, O. H., Smith, N. D., & Warnick, W. L. (1988, April). *Target acquisition and analysis training system: An exploratory investigation of vehicle identification performance with black hot and white hot thermal images* (Research Note 88-34). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A195 725)

Jane's Information Group. (1993a). *Aircraft recognition training: Reference database* [CD-ROM]. Alexandria, VA: Author.

Jane's Information Group. (1993b). *Aircraft recognition training: Test program* [CD-ROM]. Alexandria, VA: Author.

Jane's Information Group. (1997). *Jane's defence equipment library.* [CD-ROM]. Alexandria, VA: Author.

Joint Chiefs of Staff. (1987, June). *Department of Defense dictionary of military and associated terms* (JCS Pub. 1). Washington, DC: Author.

Lintern, G. (1989). The learning strategies program: Concluding remarks. *Acta Psychologica, 71,* 301-309.

Lupart, J. L. (1995). Exceptional learners and teaching for transfer. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp, 215-228). Mahwah, NJ: Lawrence Erlbaum.

Marini, A., & Genereux, R. (1995). The challenge of teaching for transfer. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp. 1-19). Mahwah, NJ: Lawrence Erlbaum.

Nesbit, J. C., & Yamamoto, N. (1991). Sequencing confusable items in paired-associate drill. *Journal of Computer-Based Instruction, 18*(1), 7-13.

O'Kane, B. L., Biederman, I., Cooper, E. E., & Nystrom, B. (1997). An account of object identification confusions. *Journal of Experimental Psychology: Applied, 3,* 21-41.

Orentas, R. P., Zegel, F. H., & Gonzalez, J. (1991). *Desert Shield IR Guide II.* Ft. Belvoir, VA: Center for Night Vision and Electro-Optics.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review, 56,* 132-143.

Palmer, J. E., D'Agostino, J., & Lillie, T. J. (1982). *Infrared recognition and target handbook.* Ft. Belvoir, VA: Night Vision and Electro-Optics Laboratory.

Proctor, R. W. & Dutta, A. (1995). *Skill acquisition and human performance.* Thousand Oaks, CA: Sage.

Rollier, R. L., Champion, D. F., Roberson, P. R., & Graber, J. G. (1988). *Procedures and guidelines for thermal imaging* (ARI RP 88-39). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. *International Journal of Educational Research, 11*, 623-637.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 18*, 42-50.

Smith, N. D., Heuckeroth, O. H., Warnick, W. L., & Essig, S. S. (1989, January). *Target acquisition and analysis training system: Effects on combat vehicle identification (CVI) performance of number of vehicles trained, training frequency, and soldier trainability* (Research Note 89-02). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A206 500)

Smith, N. D., Heuckeroth, O. H., Warnick, W. L., & Essig, S. S. (1980, November). *Evaluation of a new approach to target acquisition training: The combat vehicle identification (CVI) training program* (Research Report 1304). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A111 732)

Smith, N. D., Shope, G. L., Heuckeroth, O. H., Warnick, W. L., & Essig, S. S. (1987). *Target acquisition and analysis training system: Evaluation of the basic thermal combat vehicle identification (TCVI) training program* (Research Report 1378). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A192 488)

Test and Experimentation Command. (1995, September). *Improved target acquisition system (ITAS): Limited user test* (Test Report 95-LU-1694A). Fort Hood, TX: Close Combat Test Directorate. (DTIC No. AD-B204 291)

U.S. Army Air Defense Artillery School. (1993, June). *Visual aircraft recognition* [CD-ROM 441-173]. Ft. Bliss, TX: Author.

Warnick, W. L., & Kubala, A. L. (1979). *Studies in long range target identification* (Research Report 1216). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A073 860)

Warnick, W. L., & Smith, N. D. (1989). *Summary of research on combat vehicle identification completed at ARI Fort Hood Field Unit, Fort Hood, Texas, 1980-1985* (Research Report 1540). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD-A217 451).

Whitmore, P. G., Cox, J. A., & Friel, D. J. (1968). *A classroom method of training aircraft recognition* (HumRRO Technical Report 68-1). Washington, DC: George Washington University, Human Resources Research Office.

APPENDIX A

DEMOGRAPHIC QUESTIONS AND PRE-POSTTESTS

DEMOGRAPHIC INFORMATION

CVI*Plus* Version 1.3B (Experiments 2 and 3)

The following questions were answered by each soldier during sign-on procedures to CVI*Plus*.

1. Is this the first time using CVI*Plus* on this computer?
   Yes		No

2. Subject Group (if not applicable, type "na" and click Next)
   This field was used to assign soldiers to experimental conditions.

3. Last four digits of SSN ________

4. Last Name ________

5. First Name ________

6. Age ________

7. Military Service

   Army
   Marines

8. Your Rank Is

General	Command Sergeant Major
Lieutenant General	Sergeant Major
Major General	Master Sergeant
Brigadier General	First Sergeant
Colonel	Sergeant First Class
Lieutenant Colonel	Staff Sergeant
Major	Sergeant
Captain	Specialist
First Lieutenant	Corporal
Second Lieutenant	Private First Class
Master Warrant Officer	Private
Chief Warrant Officer (CW2, 3, or 4)	Civilian Department of Army
Warrant Officer	Civilian

9. Current Duty Station:

29 Palms	Fort Hood
Aberdeen Proving Ground	Fort Knox
Camp Lejeune	Fort Lewis
Camp Pendleton	Fort Riley
Fort Benning	Fort Rucker
Fort Belvoir	Fort Sill
Fort Bliss	Fort Stewart
Fort Bragg	National Training Center
Fort Carson	Other

10. MOS (Example: 11K8D) __________

11. Years in the Military

Less than 6 months
6 months to 1 year
1-3 years
4-6 years
7-10 years
11-15 years
16-20 years
Over 20 years

12. Months at current duty position

Less than 6 months
6-12 months
13-18 months
19-24 months
25-36 months
37-48 months
Over 48 months

13. Which thermal sights have you trained on for at least 5 hours? Select one or more.

M1 Thermal Imaging System	AH-1 Cobra PNVS
Bradley ISU/TOW	OH-58 Kiowa Mast Mounted Sight
TOW (AN/TAS-4)	AH-64 Apache TADS
M60 Tank Thermal Sight	Javelin CLU
LAV-25 (USMC)	Dragon
Thermal Weapon Sight (TWS)	Other

14. Have you used this sight in a battlefield situation?

    Yes   No

    If Yes: In which tactical situation?

        Vietnam
        Grenada or Panama
        Desert Storm Conflict
        Somalia
        Bosnia
        Sinai or Saudi Arabia
        NTC, JRTC or other Training Center
        Other

15. Duty phone number - In case we need to contact you again:

    Comm: __________

    DSN: __________

Information on duty position was obtained through interviews with each soldier.

CVI*Plus* Version 1.0 (Experiment 1)

Questions 1, 3, 4, 5, 7, 8, 9, 10, 11, and 12 were presented during the sign-on procedures for Version 1.0. However, rank, duty station, years in military, and months at duty position were fill-in-the blank items, not multiple-choice. In addition Version 1.0 required each individual to indicate current duty position, and soldiers were asked about their prior experience with thermal sights. No information on age, combat experience, Combat Training Center experience was obtained.

Table A-1

Vehicles on the FLIR and Visible Pretests and Posttests by Aspect Angle

	Aspect Angle for FLIR and Visible Images							
	Flanks		Front Aspects			Rear Aspects		
Vehicle	Left	Right	Front	Left Front	Right Front	Rear	Left Rear	Right Rear
M1	X		X					X
M2	X		X					X
T72	X				X		X	
BTR	X				X	X		
BMP		X				X - Vis	X	X - FLIR
LAV		X		X		X		
M60		X		X		X		
M113	X			X				X
ZSU		X	X				X	
HMMWV		X	X				X	
Total	5	5	4	3	2	3 (4)	4	3 (4)

Note. Each vehicle was presented 3 times. For 9 of the 10 vehicles one flank view, one front view, and one rear view were presented. For the BMP, this pattern differed slightly, as shown in the table.

APPENDIX B

EXPERIMENT 1

PART TASK TRAINING

Tables

Figures

Table B-1

Experiment 1: Distribution of Training and Testing Times (in min:sec) as a Function of FLIR Training

Training and Testing Events	Training Condition							
	Diad (n = 12)				Triad (n = 11)			
	M	*SD*	Min	Max	*M*	*SD*	Min	Max
Pre-FLIR	5:47	1:14	4:11	9:06	5:10	1:39	2:46	7:44
Pre-Vis	3:04	0:45	2:04	4:25	3:25	1:24	1:56	6:18
SC Exercises	34:23	11:41	20:44	56:07	49:49	12:36	31:38	70:17
Post-FLIR	4:01	0:43	2:46	4:56	4:47	2:00	1:39	9:29
Post-Vis	3:45	0:46	2:26	4:57	4:36	2:05	2:13	7:53

Note. Times for transfer test not included because of the limited number of soldiers who took it.

Table B-2

Experiment 1: Mean Number of Sessions to Meet Criterion on Signature Challenge Part-Task Vehicle Sets

	Diad (n = 12)			Triad (n = 12)		
Vehicle Set	*M*	Min	Max	*M*	Min	Max
M1 - T72	2.08	1	4	-----	-----	-----
M60 - M551	3.50	1	9	-----	-----	-----
M2 - ZSU	1.58	1	3	-----	-----	-----
BTR- LAV	3.33	2	7	2.92	1	8
M1-T72- M60	-----	-----	-----	4.25	2	10
M2 - M551 - ZSU	-----	-----	-----	5.41	1	9

Table B-3

Experiment 1: Mean Vehicle Identification Scores (% correct) and Response Times (sec) on First Session of Signature Challenge Part- and Whole-Task Vehicle Sets

	Part-Task Vehicle Sets				Whole-Task Vehicle Set					
	Diad		Triad		Diad		Triad		Both	
Vehicle	Score	Time	Score	Time	Score	Time	Score	Time	Score	Time
M1	85	1.92	65	2.18	79	2.70	77	3.33	78	3.00
T72	73	2.06	74	2.03	67	2.95	62	3.40	65	3.17
M60	64	1.83	53	2.18	69	2.70	78	3.31	73	2.99
M551	65	1.80	55	2.34	39	3.41	54	3.98	47	3.68
M2	91	1.51	62	2.07	77	2.79	73	3.20	75	2.99
ZSU	72	1.74	57	2.39	42	3.51	47	3.98	44	3.74
LAV	69	1.84	70	2.26	71	2.72	52	3.81	62	3.24
BTR	58	1.78	71	2.16	78	2.71	57	3.58	68	3.12

Note. Means based on 10 trials per vehicle per soldier in part-task sets; 12 soldiers in Diad and Triad conditions. Means based on 8 trials per vehicle per soldier in whole-task set; 12 soldiers in Diad and 11 soldiers in Triad.

Table B-4

Experiment 1: Mean Vehicle Identification Scores (% correct) on the Vehicle ID Test by Transfer Condition

Transfer Condition	Subscores	Total Score
No Transfer: Range 1, Night, White-Hot	Front- 81% Rear - 69%	75%
Transfer to Greater Range: Range 2, Night, White-Hot	Left Flank - 66% Left Rear Oblique - 69% Right Front Oblique - 59%	65%
Transfer to Day Thermal: Range 1, Day, White-Hot	Right Flank - 59% Right Rear Oblique - 78% Left Front Oblique - 53%	63%

Note. Means for each aspect based on 24 trials; 8 vehicles and 4 soldiers.

Table B-5

Experiment 1: Mean Identification Scores for Each Vehicle on the Vehicle ID Transfer Test and the Aspects Associated with Low Identification Scores

Vehicle	% Correct	Difficult Aspects[a]	Transfer Condition
M60	81	Left flank	Range
M1	78	Left front oblique Right flank	Day Day
BTR	75	Rear Right flank	None Day
T72	69	Right flank Right front oblique	Day Range
M2	66	Front Left flank Right rear oblique	None Range Day
M551	62	Left front oblique Left rear oblique Right front oblique	Day Range Range
LAV	62	Left front oblique Left flank Left rear oblique Rear	Day Range Range None
ZSU	41	Left front oblique Left flank Left rear oblique Right rear oblique Right flank Right front oblique	Day Range Range Day Day Range

[a] Aspect indicated if vehicle identified correctly no more than 50% of the time when aspect displayed. Means based on 8 trials per soldier; 4 soldiers.

Table B-6
Experiment 1: Mean Aspect Scores (% correct) on the Vehicle ID Transfer Test

Aspect Angle	Aspect Score	Transfer Condition
Left Flank	84%	Range
Right Flank	81%	Day
Left Front Oblique	72%	Day
Front	72%	None
Right Rear Oblique	69%	Day
Left Rear Oblique	62%	Range
Rear	62%	None
Right Front Oblique	19%	Range

Note. Each mean based on 24 trials; 8 vehicles and 4 soldiers.

In the vehicle confusion matrix presented in Table B-7, the diagonal cells contain % correct responses. The off-diagonal cells represent the percentage of instances where the vehicle presented was confused with (misidentified as) another vehicle. The columns, by row, sum to 100%, within rounding error. In the pretest and posttests, each vehicle was displayed 3 times; each presentation was a different aspect angle. The category of "other" refer to the M35 2.5 ton truck, M814 5 ton truck, and the ZIL-131 truck. N for each test and for each vehicle presented (column) was 51 (3 trials or displays per soldier; 17 soldiers).

Table B-7

Experiment 1: Vehicle Confusion Matrix from Pretests and Posttests

Vehicle Response	M1	T72	M60	M2	BMP	LAV	BTR	M113	ZSU	HMMWV
	Vehicle Presented									
M1 PreVis	**78**	25	10	8	8	0	2	0	14	0
PreIR	**92**	12	2	2	8	0	2	0	14	0
PostVis	**51**	18	4	4	4	4	0	0	18	0
PostIR	**82**	14	0	8	2	0	0	0	16	0
T72 PreVis	10	**44**	10	0	12	0	4	0	2	0
PreIR	8	**68**	0	6	6	0	0	0	4	0
PostVis	10	**41**	6	10	10	2	4	0	8	0
PostIR	6	**80**	2	0	8	0	0	0	2	0
M60 PreVis	2	6	**42**	14	8	6	2	4	20	0
PreIR	0	8	**80**	6	0	0	2	0	10	0
PostVis	10	10	**20**	8	4	4	2	6	2	0
PostIR	6	4	**90**	2	2	0	0	0	4	0
M2 PreVis	6	2	4	**49**	12	4	4	0	16	0
PreIR	0	0	0	**74**	12	12	0	12	14	0
PostVis	10	12	6	**45**	8	10	0	4	6	0
PostIR	0	0	0	**84**	18	14	0	20	6	0
BMP PreVis	0	0	0	4	**19**	8	20	10	0	0
PreIR	0	0	0	0	**24**	2	2	6	2	0
PostVis	0	4	0	2	**16**	12	20	16	12	0
PostIR	0	0	0	0	**20**	2	0	8	0	0
LAV PreVis	0	0	4	4	6	**31**	25	6	2	0
PreIR	0	2	0	2	2	**59**	14	0	6	0
PostVis	0	0	4	2	8	**29**	22	12	8	0
PostIR	0	0	0	0	8	**69**	6	0	2	0
BTR PreVis	0	0	6	4	4	24	**27**	6	2	0
PreIR	0	0	0	0	8	22	**74**	0	0	0
PostVis	2	2	8	8	10	18	**24**	2	10	0
PostIR	0	0	0	2	0	8	**94**	0	0	0
M113 PreVis	0	4	0	6	2	2	0	**60**	10	0
PreIR	0	0	0	4	6	0	0	**66**	0	0
PostVis	4	0	10	6	2	4	4	**49**	2	0
PostIR	0	0	0	2	2	0	0	**61**	2	0
ZSU PreVis	0	2	0	4	12	12	12	2	**18**	0
PreIR	0	0	0	0	14	6	2	6	**44**	0
PostVis	2	6	2	6	10	4	10	4	**16**	0
PostIR	0	0	0	2	20	6	0	6	**65**	2
HMMWV										
PreVis	0	0	0	0	0	0	0	0	0	**84**
PreIR	0	0	0	0	0	0	0	0	0	**90**
PostVis	0	0	0	0	0	0	2	0	0	**69**
PostIR	0	0	0	0	0	0	0	0	0	**76**
M551 PreVis	4	15	20	0	4	4	2	6	4	0
PreIR	0	10	18	6	10	0	4	0	6	0
PostVis	10	8	37	6	10	0	0	4	14	0
PostIR	6	2	8	0	2	2	0	0	2	0
Other PreVis	0	2	4	6	10	10	2	6	12	16
PreIR	0	0	0	0	8	0	0	10	0	10
PostVis	2	0	4	4	19	13	14	4	6	22
PostIR	0	0	0	0	20	0	0	6	2	22

Table B-8

Experiment 1: Pre-Post Correlations with First Session Scores and Number of Sessions to Criterion on Signature Challenge

Signature Challenge Vehicle Set	Pre Visible	Pre FLIR	Post Visible	Post FLIR
Scores on First Session for Each Vehicle Set				
M1-T72	.29	.52	.42	.03
M60-M551	.46	.70**	.60**	.50
M2-ZSU	-.05	.02	-.01	.09
BTR-LAV	.04	.06	.09	.11
M1-T72-M60	.32	.42	.63*	.78**
M2-M551-ZSU	.58*	.60**	.23	.31
Final Vehicle Set	.58*	.62**	.44*	.22
Number of Sessions to Criterion for Each Vehicle Set				
M1-T72	-.12	-.30	-.27	-.24
M60-M551	-.17	-.47	-.34	-.23
M2-ZSU	.02	.15	.09	.16
BTR-LAV	.04	.03	-.12	-.02
M1-T72-M60	.06	-.10	-.46	-.61**
M2-M551-ZSU	-.61**	-.48	-.60	-.61**

*$p < .05$ **$p < .01$

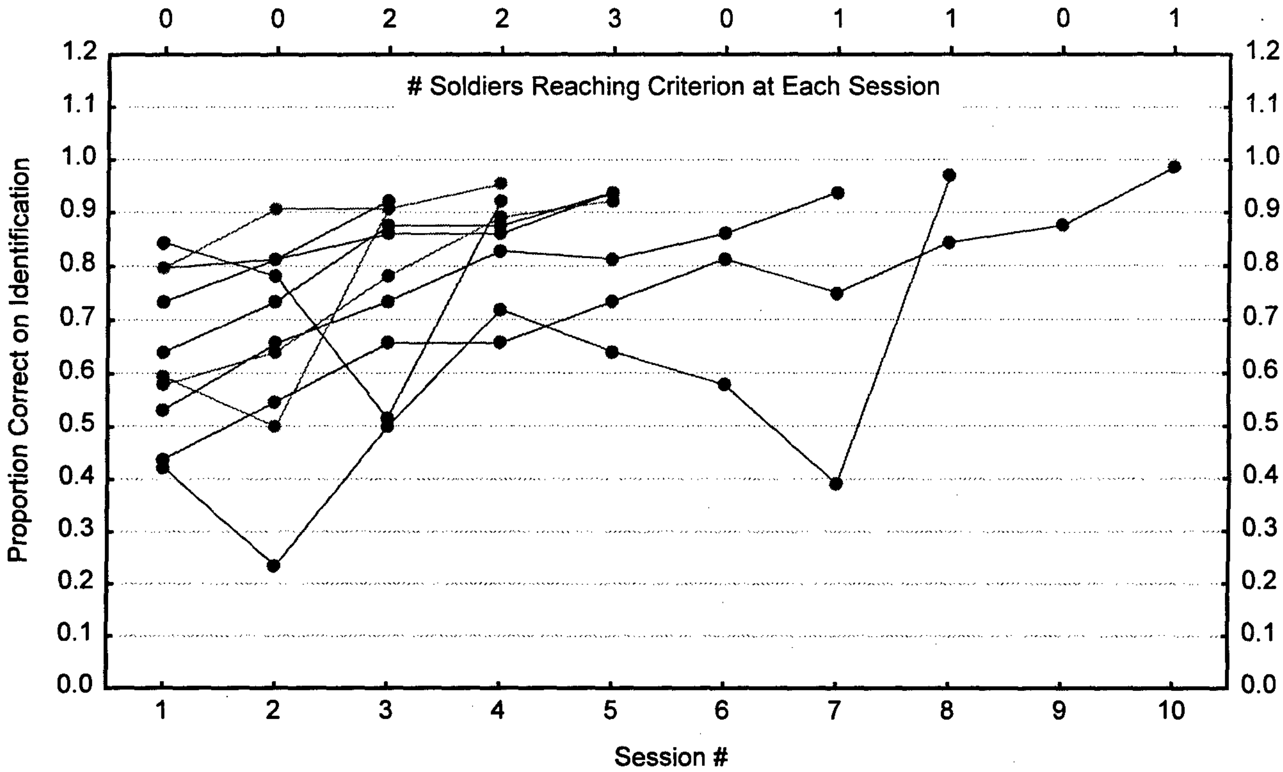

Figure B-1. Experiment 1: Learning curves for soldiers who met the criterion of 90% on the whole-task vehicle set in Signature Challenge.

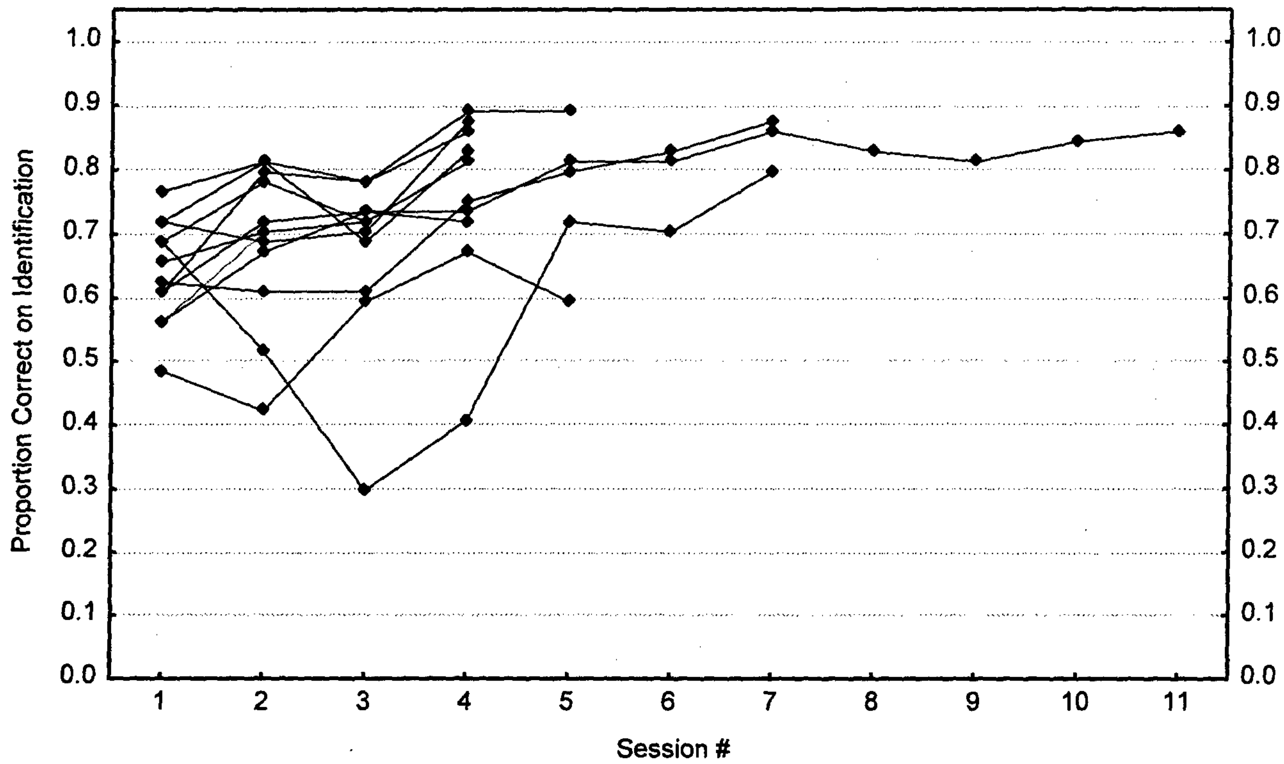

Figure B-2. Experiment 1: Learning curves for soldiers who did not meet the criterion of 90% on the whole-task vehicle set in Signature Challenge.

APPENDIX C

EXPERIMENT 2

FIXED-PACED WITH KNOWLEDGE OF RESULTS VERSUS SELF-PACED WITH ENHANCED KNOWLEDGE OF RESULTS

Tables

Figures

Table C-1

Experiment 2: Distribution of Training and Testing Times (min:sec) as a Function of FLIR Training

Training and Testing Events	Training Condition							
	Self-KR+ (n = 16)				Fixed-KR (n = 18)			
	M	*SD*	Min	Max	*M*	*SD*	Min	Max
Pre-FLIR	9:11	3:38	3:50	16:27	8:48	3:03	5:47	15:21
Pre-Vis	4:29	1:27	2:02	7:23	5:01	1:33	2:29	7:58
Veh ID	59:41	23:30	22:35	118:55				
SC					54:47	18:40	15:59	90:50
Transfer	15:15	6:40	7:05	30:31	13:06	4:03	7:39	21:27
Post-FLIR	5:23	2:14	2:26	9:19	5:00	2:09	2:14	11:39
Post-Vis	4:59	2:20	2:03	10:39	4:58	2:12	1:57	9:13

Table C-2

Experiment 2: Time to Respond (sec) to Vehicles in the All-Vehicle Set in Self-KR+ and Fixed-KR Training - Entire sample

	Self-KR+		Fixed-KR	
Vehicle	*M*	*SD*	*M*	*SD*
M1	29.78	33.97	2.75	1.49
M60	18.30	18.48	2.19	1.11
T72	29.81	35.12	3.08	1.47
M2	31.09	37.54	2.30	1.38
BTR	27.69	25.39	2.52	1.45
LAV	33.29	30.01	3.08	1.61
Total	28.33	31.05	2.65	1.46

Note. Means and standard deviations based on trial data; 128 trials per vehicle (8 displays; 16 soldiers in each condition).

Table C-3

Experiment 2: Response Profile for Soldiers who Required Repeated Training Sessions to Achieve Criterion on the All-Vehicle Set

Soldier #	Response Profile						
	# Sessions			Pretest FLIR Score on Vehicles Trained	Score on First Session of All-Vehicle Set	Response Time (sec)	
	All	91%	85%			First Session	All Sessions
	Self-KR+ Training Condition						
Self-KR+ #1	4	---	3	89	75	36.20	31.01
Self-KR+ #2	4	2	3	94	69	20.52	13.61
Self-KR+ #3	4	3	1	72	85	19.72	16.00
Self-KR+ #4	5	---	1	39	87	14.41	10.32
Self-KR+ #5	7	---	7	17	40	15.75	11.29
Mean	4.80	2.5	3.00	62	71	21.32	16.45
	Fixed-KR Training Condition						
Fixed-KR #1	4	4	3	94	75	3.33	2.74
Fixed-KR #2	5	5	4	67	65	2.59	2.31
Fixed-KR #3	8	8	1	50	87	1.79	2.99
Fixed-KR #4	9	9	6	83	79	3.00	2.08
Fixed-KR #5	19	19	3	72	77	2.55	1.63
Mean	7.00	7.00	3.40	73	77	2.65	2.35

Note. The close-up range was used in the FLIR pretest; Range 2 was used in training.

Table C-4

Experiment 2: Aspect Angle Confusions in Self-KR+ Training - First Session

Aspect	Mean % Correct	Confusions (> 10%)
Left flank	96	----
Right flank	94	----
Rear	84	Front (16%)
Left rear oblique	81	----
Right rear oblique	75	----
Left front oblique	59	Left flank (16%)
Right front oblique	57	Right rear oblique (17%); Right flank (22%)
Front	57	Rear (38%)
Mean	75	

Note. Aspect means based on 96 trials - 6 vehicles and 16 soldiers.

Table C-5

Experiment 2: Vehicle Confusions by Training Condition and Soldier Sample on Transfer Test

Vehicle Presented	Vehicle Response					
	M1	M60	T72	M2	BTR	LAV
	Entire Sample					
M1 Self-KR+	**82**	7	8	1	2	1
Fixed-KR	**73**	8	14	4	0	1
M60 Self-KR+	8	**73**	13	8	2	2
Fixed-KR	3	**84**	8	1	0	1
T72 Self-KR+	18	5	**74**	1	0	2
Fixed-KR	12	7	**73**	4	1	0
M2 Self-KR+	5	2	3	**80**	5	5
Fixed-KR	2	7	4	**73**	4	3
BTR Self-KR+	1	1	4	4	**73**	18
Fixed-KR	1	3	5	3	**71**	14
LAV Self-KR+	2	1	5	8	13	**73**
Fixed-KR	1	4	4	4	11	**73**
	Pass85 Sample					
M1 Self-KR+	**95**	3	3	0	0	0
Fixed-KR	**77**	7	13	3	0	0
M60 Self-KR+	4	**81**	14	0	0	1
Fixed-KR	3	**88**	6	1	0	1
T72 Self-KR+	9	3	**89**	0	0	0
Fixed-KR	11	6	**78**	2	1	0
M2 Self-KR+	1	1	1	**90**	0	6
Fixed-KR	1	4	3	**88**	0	1
BTR Self-KR+	1	0	3	3	**85**	9
Fixed-KR	1	1	3	4	**80**	9
LAV Self-KR+	0	1	1	6	8	**84**
Fixed-KR	0	4	4	4	12	**76**

Note. Percent correct responses are in bold on the diagonal; confusions, on the off-diagonal. Number of trials (i.e., displays) per vehicle was six. Number of soldiers for entire sample: Self-KR+ = 16, Fixed-KR = 17. Number of soldiers for Pass85 sample: Self-KR+ = 10, Fixed-KR = 14.

Table C-6

Experiment 2: Aspect Angle Confusions on Vehicle ID Test

Aspect Presented	Mean % Correct		Confusions (>10%)	
	Self-KR+	Fixed-KR	Self-KR+	Fixed-KR
	Entire Sample			
No Transfer Condition				
Front	73	72	Rear - 22%	Rear - 23%
Rear	89	75	Front - 10%	Front - 25%
Transfer Conditions				
Left flank	91	88	----	----
Right flank	88	77	Right rear - 11%	Right rear - 16%
Left rear oblique	85	63	----	Right front - 13%
Right rear oblique	80	64	----	Rear - 12%
Left front oblique	70	69	Right rear - 10%	Right rear - 11%
Right front oblique	55	44	Right rear - 23%	Right rear - 27% Right flank - 14%
	Pass85 Sample			
No Transfer Condition				
Front	85	79	Rear - 12%	Rear - 17%
Rear	87	74	Front - 12%	Front - 25%
Transfer Condition				
Left flank	97	92	----	----
Right flank	87	80	Right rear - 12%	Right rear - 13%
Left rear oblique	92	70	----	Right front - 11%
Right rear oblique	87	68	----	Rear - 10%
Left front oblique	85	74	----	Right rear - 11%
Right front oblique	63	45	Right rear - 15%	Right rear - 26% Right front - 12%

Note. Number of trials (i.e., displays) per aspect was six. Number of soldiers for entire sample: Self-KR+ = 16, Fixed-KR = 17. Number of soldiers for Pass85 sample: Self-KR+ = 10, Fixed-KR = 14.

Table C-7

Experiment 2: Time to Respond (sec) on the Transfer Test for Self-KR+ and Fixed-KR Groups as a Function of Vehicle and Aspect Angle - Entire Sample

	Self-KR+		Fixed-KR	
Vehicle	*M*	*SD*	*M*	*SD*
M1	17.82	19.78	15.14	14.87
M2	22.51	30.44	19.61	23.32
T60	15.35	16.15	13.05	9.47
T72	17.78	21.37	18.42	18.67
BTR	19.60	23.53	16.04	14.63
LAV	23.04	22.30	18.66	13.23
Aspect				
No Transfer				
Front	21.41	28.50	16.62	15.27
Rear	17.98	21.53	15.88	16.20
Transfer				
Left flank	15.68	14.39	13.65	8.33
Right flank	14.34	17.20	13.69	18.63
Left rear oblique	17.00	17.19	18.18	23.15
Right rear oblique	21.17	29.76	21.72	19.16
Left front oblique	23.05	21.66	17.05	10.46
Right front oblique	25.49	25.45	17.76	14.08
Total	19.52	22.76	16.82	16.40

Note. Means and standard deviations based on trial data.
Self-KR+: Vehicle = 128 trials. Aspect = 96 trials
Fixed-KR: Vehicle = 136 trials. Aspect = 102 trials.

Table C-8

Experiment 2: Time to Respond (sec) on the Transfer Test for Self-KR+ and Fixed-KR Groups as a Function of Vehicle and Aspect Angle - Pass85 Sample

	Self-KR+		Fixed-KR	
Vehicle	*M*	*SD*	*M*	*SD*
M1	9.81	6.81	14.74	15.29
M2	18.67	32.58	16.30	15.89
M60	11.06	11.24	11.55	8.35
T72	13.38	18.21	16.33	13.41
BTR	15.30	20.37	15.41	12.78
LAV	17.93	17.63	17.77	11.94
Aspect				
No Transfer				
Front	17.08	29.36	15.35	14.64
Rear	11.30	8.66	15.26	16.68
Transfer				
Left flank	11.66	11.80	13.02	8.37
Right flank	9.75	6.01	11.54	7.25
Left rear oblique	11.23	8.57	13.47	8.75
Right rear oblique	15.91	30.35	20.34	18.86
Left front oblique	16.05	16.03	15.63	9.25
Right front oblique	21.90	25.31	18.19	15.27
Total	14.36	19.72	15.35	13.27

Note. Means and standard deviations based on trial data.
Self-KR+: Vehicle = 80 trials. Aspect = 60 trials
Fixed-KR: Vehicle = 112 trials. Aspect = 84 trials.

Table C-9

Experiment 2: Mean Percentage of Vehicles Identified Correctly on the Pretests and Posttests as a Function of FLIR Training and Type of Imagery for Entire Sample

Vehicle	Visible Imagery			FLIR Imagery		
	Pretest	Posttest	Difference: Post-Pre	Pretest	Posttest	Difference Post-Pre
Vehicles with FLIR Training						
M60A3	60	90	30	51	92	41
M2/M3	79	92	13	71	88	17
M1	80	87	7	61	81	20
BTR-80	42	72	30	47	81	34
T72	64	83	19	57	73	16
LAV-25	50	76	26	40	64	24
Vehicles without FLIR Training						
HMMWV	95	96	1	87	97	10
M113	93	97	4	92	94	2
BMP-2	45	42	-3	35	34	-1
ZSU 23-4	37	35	-2	28	27	-1
Trained	62	83	21	55	80	25
Not Trained	67	67	0	61	63	2
Overall[ab]	65	77	12	57	73	16

Note. Vehicle means based on 3 trials (displays) per soldier; 33 soldiers.
Significant main effects for Pre-post and Visible-FLIR factors.
Pre-Post: $F(1, 31) = 50.99$, $p < .0000$
Vis-FLIR: $F(1, 31) = 21.32, p < .0000$
[a] Pre-post x Trn-NoTrn: $F(1, 31) = 31.91, p < .0000$
[b] Vehicles weighed equally in overall means.

Table C-10

Experiment 2: Vehicle Scores on Pretests and Posttests as a Function of Training Condition

Vehicle	Visible Imagery			FLIR Imagery		
	Pretest	Posttest	Difference: Post-Pre	Pretest	Posttest	Difference Post-Pre
Self-KR+ Training						
Entire Sample ($n = 16$)	66	77	11	56	75	19
Pass85 Sample[a] ($n = 10$)	80	87	7	70	88	18
Fixed-KR Training						
Entire Sample ($n = 17$)	63	74	11	57	68	11
Pass85 Sample[a] ($n = 10$)	70	79	9	64	71	5

Note. Train and no train conditions weighed equally in means.
[a] Tng Cond x pre-post x Vis-FLIR: $F(1, 22) = 5.16, p < .0337$

In the vehicle confusion matrices presented in Tables C-11 and C-12, the diagonal cells contain the percent correct responses. The off-diagonal cells represent the percentage of instances where the vehicle presented was confused with, or misidentified as, another vehicle. The columns sum to 100%, within rounding error. In the pretests and posttests, each vehicle was displayed 3 times; each presentation was a different aspect angle. The category of "other" refers to the M35 2 ton truck, M814 5 ton truck, and the ZIL-131 truck.

Number of trials for entire sample, Table C-11: Pretests and PostIR = 102; PostVis = 99.

Number of trials for Pass85 sample, Table C-12: Pretests and PostIR = 72; PostVis = 69.

Table C-11

Experiment 2: Vehicle Confusion Matrix from Pretests and Posttests – Entire Sample

Vehicle Response	Vehicle Presented									
	M1	T72	M60	M2	BMP	LAV	BTR	M113	ZSU	HMMWV
M1 PreVis	**80**	6	5	7	7	1	1	2	9	0
PostVis	**87**	8	2	3	10	2	1	0	14	0
PreIR	**62**	13	5	4	9	1	1	0	10	0
PostIR	**81**	11	1	4	4	1	0	1	16	0
T72 PreVis	8	**64**	15	1	11	0	1	0	6	0
PostVis	6	**83**	4	0	19	0	0	0	14	0
PreIR	12	**57**	11	1	8	2	2	1	8	0
PostIR	4	**73**	4	0	12	2	1	0	24	0
M60 PreVis	0	8	**60**	4	9	1	2	0	9	0
PostVis	4	4	**90**	2	3	2	3	0	6	0
PreIR	8	9	**51**	6	9	2	1	1	13	0
PostIR	6	10	**92**	2	15	1	1	0	10	0
M2 PreVis	1	1	2	**79**	7	4	1	1	12	2
PostVis	2	0	3	**92**	7	4	0	0	13	3
PreIR	8	3	4	**72**	28	14	1	1	9	7
PostIR	8	5	1	**88**	33	14	2	0	13	2
BMP PreVis	0	1	3	0	**45**	18	27	0	2	0
PostVis	0	3	1	0	**42**	0	3	0	5	0
PreIR	2	1	6	2	**35**	16	25	1	6	0
PostIR	0	1	0	1	**34**	1	2	0	2	0
LAV PreVis	0	6	4	0	2	**50**	10	0	6	0
PostVis	0	0	0	0	3	**76**	20	0	2	0
PreIR	1	1	4	2	0	**40**	8	0	3	3
PostIR	0	0	0	1	0	**64**	11	0	1	0
BTR PreVis	2	4	3	1	4	14	**42**	0	6	0
PostVis	0	0	0	0	6	15	**72**	0	0	0
PreIR	1	2	7	2	1	14	**47**	1	8	1
PostIR	0	0	1	1	1	18	**81**	1	1	0
M113 PreVis	0	1	0	2	0	2	1	**93**	2	0
PostVis	0	0	0	1	0	0	0	**97**	0	0
PreIR	0	1	0	4	0	2	1	**92**	2	0
PostIR	1	0	0	1	0	0	1	**94**	2	0
ZSU PreVis	1	2	0	3	6	7	5	1	**37**	0
PostVis	0	0	0	0	3	1	1	1	**35**	0
PreIR	2	4	1	2	3	4	4	1	**28**	0
PostIR	0	0	0	0	0	0	0	1	**27**	0
HMMWV										
PreVis	0	0	0	0	0	0	0	2	0	**95**
PostVis	0	0	0	0	0	0	0	0	0	**96**
PreIR	0	1	0	0	0	1	3	0	0	**87**
PostIR	0	0	0	0	0	0	1	1	0	**97**
M551 PreVis	6	7	8	2	5	1	5	0	9	0
PostVis	1	2	0	2	3	0	0	1	8	0
PreIR	4	8	8	4	4	3	4	1	9	0
PostIR	0	0	0	1	1	2	2	2	5	0
Other PreVis	2	1	1	1	5	3	6	1	3	3
PostVis	0	0	0	0	3	0	0	1	2	1
PreIR	1	1	4	2	3	2	3	1	5	2
PostIR	0	0	0	0	0	0	0	0	0	1

Table C-12

Experiment 2: Vehicle Confusion Matrix from Pretests and Posttests – Pass85 Sample

Vehicle Response	Vehicle Presented									
	M1	T72	M60	M2	BMP	LAV	BTR	M113	ZSU	HMMWV
M1 PreVis	**92**	0	7	4	4	1	1	3	10	0
PostVis	**93**	3	0	1	6	0	1	0	14	0
PreIR	**76**	7	3	1	6	1	0	0	10	0
PostIR	**89**	8	0	0	4	0	0	1	13	0
T72 PreVis	3	**78**	7	0	8	0	1	0	6	0
PostVis	6	**94**	1	0	17	0	0	0	12	0
PreIR	8	**67**	6	1	4	3	0	1	9	0
PostIR	4	**81**	3	0	6	0	0	0	17	0
M60 PreVis	0	6	**72**	1	7	0	1	0	4	0
PostVis	1	0	**99**	0	0	0	0	0	7	0
PreIR	6	8	**69**	3	7	0	0	0	8	0
PostIR	4	7	**97**	1	14	1	0	0	8	0
M2 PreVis	1	0	0	**90**	1	4	1	1	8	3
PostVis	0	0	0	**96**	4	6	0	0	7	4
PreIR	3	0	0	**83**	28	13	0	1	6	7
PostIR	3	3	0	**94**	28	13	0	0	14	2
BMP PreVis	0	0	1	0	**61**	14	26	0	0	0
PostVis	0	3	0	0	**59**	0	4	0	6	0
PreIR	1	1	0	0	**46**	14	26	1	4	0
PostIR	0	1	0	0	**46**	0	1	0	1	0
LAV PreVis	0	6	1	0	1	**61**	10	0	1	0
PostVis	0	0	0	0	1	**87**	14	0	3	0
PreIR	1	1	4	1	0	**46**	1	0	1	1
PostIR	0	0	0	1	0	**74**	4	0	0	0
BTR PreVis	0	4	3	0	6	13	**53**	0	6	0
PostVis	0	0	0	0	7	6	**80**	0	0	0
PreIR	0	0	4	0	0	17	**63**	0	8	0
PostIR	0	0	0	0	1	13	**94**	0	1	0
M113 PreVis	0	0	0	1	0	3	0	**93**	3	0
PostVis	0	0	0	1	0	0	0	**96**	0	0
PreIR	0	1	0	3	0	1	1	**94**	3	0
PostIR	0	0	0	0	0	0	0	**94**	3	0
ZSU PreVis	0	0	0	0	4	3	1	1	**51**	0
PostVis	0	0	0	0	1	1	0	1	**43**	0
PreIR	1	3	1	1	4	3	1	0	**36**	0
PostIR	0	0	0	1	0	0	0	1	**36**	0
HMMWV										
PreVis	0	0	0	0	0	0	0	0	0	**93**
PostVis	0	0	0	0	0	0	0	0	0	**94**
PreIR	0	1	0	0	0	1	1	0	0	**92**
PostIR	0	0	0	0	0	0	0	0	0	**96**
M551 PreVis	3	6	8	1	4	0	3	0	8	0
PostVis	0	0	0	1	0	0	0	1	7	0
PreIR	3	8	8	3	6	1	3	1	7	0
PostIR	0	0	0	1	1	0	0	3	7	0
Other PreVis	1	1	0	1	3	1	1	1	3	3
PostVis	0	0	0	0	3	0	0	1	0	1
PreIR	1	1	4	2	0	0	3	0	7	0
PostIR	0	0	0	0	0	0	0	0	0	1

Table C-13

Experiment 2: Mean Response Times (sec) on Pre- and Posttests for Significant Effects

Effect	Entire Sample (n = 34)		Pass85 Sample (n = 24)	
Main Effects				
Pre-Post	Pretest	Posttest	Pretest	Posttest
	13.24	8.81	11.17	7.47
FLIR-Visible	FLIR	Visible	FLIR	Visible
	13.99	8.07	11.74	6.89
Train-NoTrain	Not Trained	Trained	Not Trained	Trained
	11.62	10.44	10.05	8.58
Interactions				
Pre-Post x FLIR-Vis	Pretest	Posttest	Pretest	Posttest
FLIR	17.78	10.19	14.88	8.60
Visible	8.71	7.43	7.45	6.33
Pre-Post x Exp Condition x Train-NoTrain	Pretest	Posttest	Pretest	Posttest
Fixed-KR -- NoTrn	14.64	8.82	13.09	7.62
Fixed-KR --Trn	11.93	8.33	10.79	7.26
Self-KR+ -- NoTrn	13.17	9.85	10.56	8.93
Self-KR+ --Trn	13.23	8.25	10.22	6.05

Note. Means based on average soldier times across trials.

Significant effects for entire sample

Pre-Post: $F(1, 31) = 56.45, p < .0000$

FLIR-Vis: $F(1, 31) = 102.85, p < .0000$

Trn-NoTrn: $F(1, 31) = 4.60, p < .0399$

Pre-Post x FLIR-Vis: $F(1, 31) = 48.83, p < .0000$

Pre-Post x Exp Cond x Trn-NoTrn: $F(1, 31) = 5.53, p < .0275$

Significant effects for the Pass85 sample

Pre-Post: $F(1, 21) = 27.64, p < .0000$

FLIR-Vis: $F(1, 21) = 72.67, p < .0000$

Trn-NoTrn: $F(1, 21) = 4.84, p < .0391$

Pre-Post x FLIR-Vis: $F(1, 21) = 29.28, p < .0000$

Pre-Post x Exp Cond x Trn-NoTrn: $F(1, 21) = 5.86, p < .0245$

Table C-14

Experiment 2: Correlations Among the Training Scores for Soldiers in the Fixed-KR Condition

	Score on First SC session		# Sessions to Pass Each SC Vehicle Set			
	APCs	All Veh	Tanks	APCs	All Veh - 91%	All Veh - 85%
Score on 1st session						
Tanks	.44	.14	-.66*	-.22	-.13	.21
APCs	...	-.32	-.28	-.42	.05	.22
All Veh		...	.24	-.28	-.53	-.70*
# Sessions to Pass						
Tanks			...	.02	-.06	-.49
APCs				...	.43	.23
All Veh - 91%					...	.46
All Veh - 85%						...

* $p < .01$

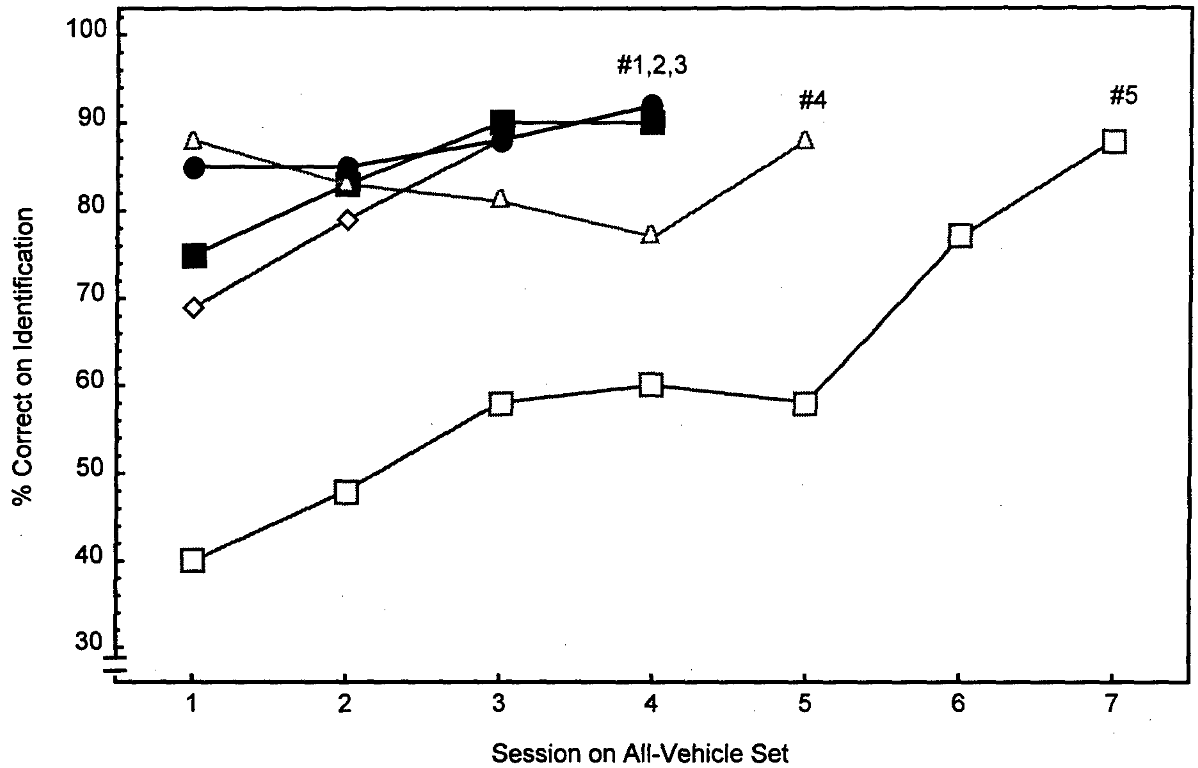

Figure C-1. Experiment 2: Learning curves for soldiers in Self-KR+ who required many sessions to achieve criterion on all-vehicle set (soldier #s refer to Table C-3).

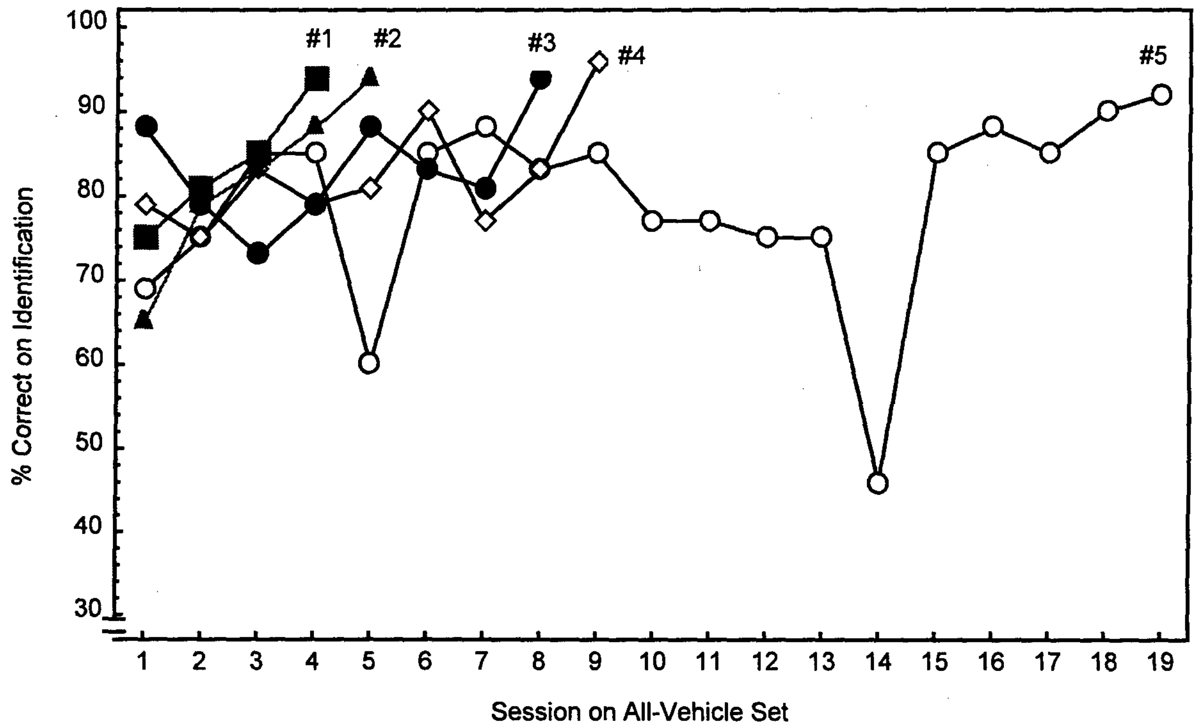

Figure C-2. Experiment 2. Learning curves for soldiers in Fixed-KR who required many sessions to achieve criterion on all-vehicle set (soldier #s refer to Table C-3).

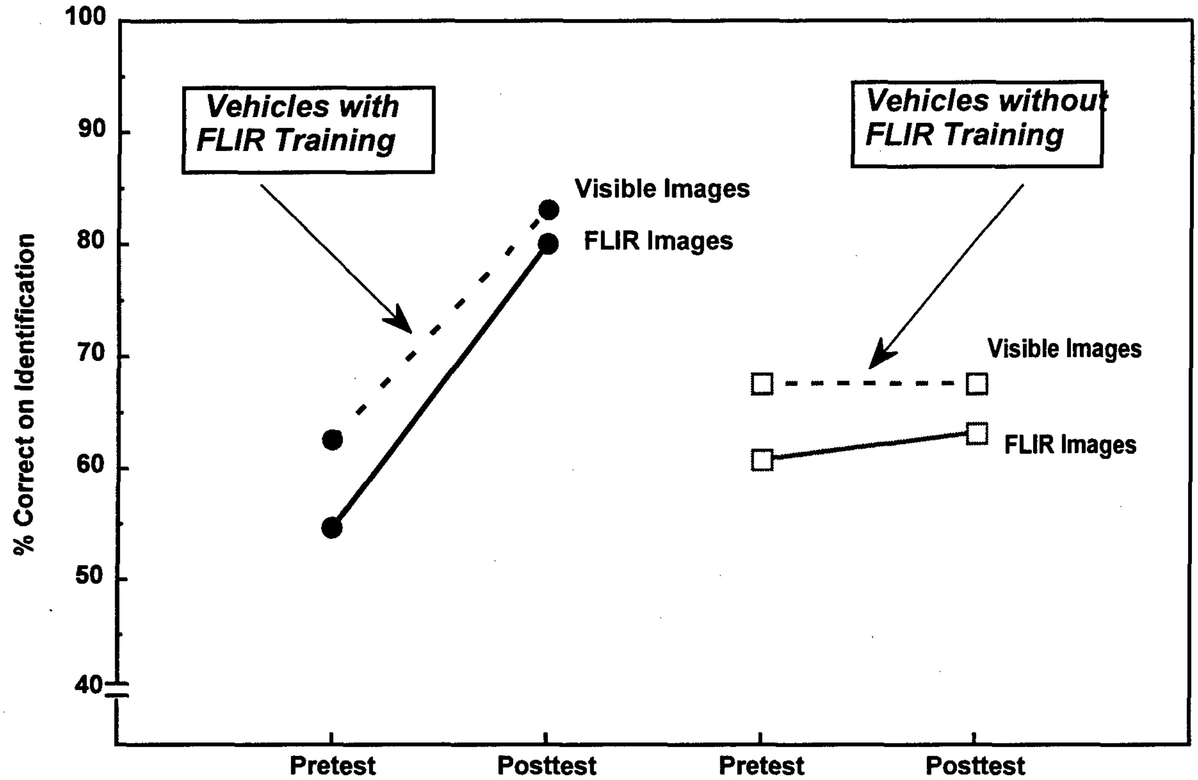

Figure C-3. Experiment2: Change in identification scores for entire sample from pretest to posttest for FLIR and visible images as a function of FLIR training.

APPENDIX D

EXPERIMENT 3

TRAINING AT NEAR AND FAR RANGES

Tables

Figures

Figures cont'd

Engine and Exhaust Test

NAME____________________ **DATE** __________

CURRENT DUTY POSITION ________________________

ENGINE LOCATION (as viewed from the driver's seat)					
Vehicle	**Where is the engine located?** [Check (✓) the block.]		Is the **engine centered**, to the **left of center**, *or* to the **right of center?** [Check (✓) the block.]		
	↓ Front	↑ Rear	Centered	Right of Center	Left of Center
M1A1					
T72					
M60A3					
M2/M3					
BMP-2					
M113					
LAV-25					
BTR-80					
HMMWV					

EXHAUST LOCATION (as viewed from the driver's seat)				
Vehicle	**Where is the exhaust located?** [Check (✓) the block(s).]			
	↓ Front	↑ Rear	→ Right Side	← Left Side
M1A1				
T72				
M60A3				
M2/M3				
BMP-2				
M113				
LAV-25				
BTR-80				
HMMWV				

Table D-1

Experiment 3: Distribution of Training and Testing Times (min:sec) as a Function of FLIR Training

Training and Testing Events	Training Condition							
	Near Practice (*n* = 33)				Far Practice (*n* = 34)			
	M	*SD*	Min	Max	*M*	*SD*	Min	Max
Pre-FLIR	8:18	2:01	4:51	12:51	7:31	1:53	3:59	12:02
Pre-Vis	4:10	1:01	2:22	5:54	3:46	0:43	2:13	5:07
Veh ID Exer	9:24	3:37	3:19	15:58	13:45	5:32	6:03	31:10
SC Exer	16:59	6:17	8:05	32:15	35:33	14:10	16:00	68:09
Transfer - SC	3:48	1:02	1:45	5:23	3:51	0:53	1:31	5:24
Transfer - ID	8:49	3:12	4:58	21:22	7:01	1:55	3:41	12:51
Post-FLIR	4:20	1:21	2:16	7:42	4:02	1:29	2:14	8:46
Post-Vis	5:02	1:11	2:46	6:58	4:34	1:17	2:51	7:33

Table D-2

Experiment 3: Mean Engine and Exhaust Location Scores (% correct) by Vehicle and Range Band of Imagery during Group Instruction

Vehicle	Dimension Tested								
	Engine Location			Symmetric Location of Engine			Exhaust Location		
		Range Band			Range Band			Range Band	
	All	Near	Far	All Ss	Near	Far	All Ss	Near	Far
In the Instruction									
M1	98	97	100	89	91	87	97	94	100
M2	96	91	100	76	71	81	89	86	94
T72	94	94	94	74	80	69	78	77	80
BMP	73	62	84	49	46	53	66	57	75
BTR	73	83	63	74	83	66	70	78	61
LAV	82	83	81	60	57	62	75	71	78
Total	86	85	87	70	71	70	79	77	81
Not in the Instruction									
M60	91	91	91	83	91	75	80	80	81
M113	98	100	97	60	51	69	82	77	87
HMMWV	98	100	97	83	86	81	68	67	69
Total	96	97	95	79	76	75	77	75	79

Note. $N = 67$; 32 for Near and 35 for Far Instruction.

Table D-3

Experiment 3: Mean Identification Scores for Significant Effects on Vehicle ID Training Exercises

Effect	Mean % Correct					
Practice[a]	Near			Far		
	78			65		
Vehicle[b]	M1	T72	M2	BMP	LAV	BTR
	79	82	83	56	62	68
Aspect[c]	Front	Rear	Left Flank		Right Flank	
	63	58	88		78	
Instruct x Vehicle[d]	M1	T72	M2	BMP	LAV	BTR
Near	75	82	83	52	67	72
Far	83	81	83	61	56	64
Aspect x Vehicle[e]	M1	T72	M2	BMP	LAV	BTR
Left Flank	90	91	91	86	79	91
Right Flank	74	86	91	50	81	84
Front	77	83	79	41	60	38
Rear	74	67	69	48	27	60
Practice x Aspect x Vehicle[f]	M1	T72	M2	BMP	LAV	BTR
-Near Ranges						
Left Flank	96	96	96	90	82	86
Right Flank	79	96	93	55	86	90
Front	96	83	90	55	69	45
Rear	76	69	83	59	24	82
-Far Ranges						
Left Flank	83	86	86	83	76	96
Right Flank	69	76	90	45	76	79
Front	58	83	69	27	52	31
Rear	72	66	55	38	31	38

Note. $N = 67$.

[a] $F(1, 54) = 14.71, p < .0003$

[b] $F(5, 270) = 16.90, p < .0000$

[c] $F(3, 162) = 44.98, p < .0000$

[d] $F(5, 270) = 2.62, p < .02$

[e] $F(15, 810) = 5.95, p < .0000$

[f] $F(15, 810) = 1.74, p < .03$

Table D-4

Experiment 3: Mean Aspect Scores for Significant Effects on Vehicle ID Training Exercises

Effect	Mean % Correct					
Practice[a]	Near			Far		
	90			78		
Vehicle[b]	M1	T72	M2	BMP	LAV	BTR
	97	87	85	81	80	77
Aspect[c]	Front	Rear	Left Flank		Right Flank	
	75	78	98		87	
Practice x Vehicle[d]	M1	T72	M2	BMP	LAV	BTR
Near	98	90	92	89	89	83
Far	97	84	77	73	71	70
Aspect x Vehicle[e]	M1	T72	M2	BMP	LAV	BTR
Left Flank	100	98	98	100	95	97
Right Flank	97	98	90	79	95	86
Front	95	66	95	69	60	66
Rear	98	86	57	76	91	60
Instruct x Aspect[f]	Front	Rear	Left Flank		Right Flank	
Near	83	76	98		88	
Far	67	81	98		86	
Practice x Aspect[g]	Front	Rear	Left Flank		Right Flank	
Near	78	87	99		98	
Far	71	69	97		76	
Pract x Aspect x Vehicle[h]	M1	T72	M2	BMP	LAV	BTR
-Near Ranges						
Left Flank	100	100	100	100	97	97
Right Flank	97	100	97	100	93	100
Front	97	66	93	72	72	69
Rear	100	97	79	83	93	72
-Far Ranges						
Left Flank	100	97	97	100	93	97
Right Flank	97	97	83	59	52	72
Front	93	66	97	66	48	62
Rear	97	76	34	70	90	48

$N = 67$.

[a] $F(1, 54) = 32.88, p < .0000$

[b] $F(5, 270) = 13.28, p < .0000$

[c] $F(3, 162) = 32.06, p < .0000$

[d] $F(5, 270) = 2.40, p < .0354$

[e] $F(15, 810) = 7.43, p < .0000$

[f] $F(3, 162) = 6.18, p < .0005$

[g] $F(3, 162) = 6.55, p < .0005$

[h] $F(15, 810) = 2.17, p < .0061$

In Table D-4, The three-way interaction among the practice, aspect, and vehicle factors showed a few exceptions to the more general pattern illustrated in Figure 3-4. For the right flank, high scores (> 90%) occurred for all vehicles at the near range, but these scores were only maintained at the far range for the M1 and T72. For the other four vehicles, aspect scores declined as a function of range. For the front, aspect scores for both ranges were similar, within 4 percentage points, for all vehicles except the LAV, where the far score was lower than the near by 24 percentage points. For the rear, aspect scores for both ranges were similar, within 6 percentage points, for three vehicles, the M1, BMP and LAV. For the other three, the T72, M2 and BTR, the far score was lower than the near by an average of 30 percentage points.

The most consistency in the scores was for the flanks, particularly at the near ranges where all vehicle scores were at least 93%. At the far ranges, the scores were also high, but dropped below 90% for the right flank for four of the six vehicles, the M2, BMP, LAV, and BTR. Aspect confusion data show that some of the confusions were with the right rear oblique and the right front oblique views of these vehicles.

Table D-5

Experiment 3: Aspect Angle Confusion Matrix on Vehicle ID Training Exercises by Range Band

Aspect	Range Band	Aspect Response			
Presented	Presented	Left Flank	Right Flank	Front	Rear
Left Flank	Near	**99**	1	0	0
	Far	**97**	2	0	0
Right Flank	Near	0	**98**	0	0
	Far	10	**76**	0	0
Front	Near	1	1	**78**	16
	Far	1	1	**72**	20
Rear	Near	0	0	11	**87**
	Far	0	1	26	**69**

Note. $N = 174$ trials for each range band. Diagonal cells contain % correct responses; aspect response was the same as the aspect presented. Off-diagonal cells represent confusions between aspect displayed and response given. Some row percentages may not sum to 100 as some soldiers indicated the aspect to be an oblique.

Table D-6

Experiment 3: Mean Identification Scores (% correct) and Response Times (sec) on the First Session of Each Vehicle Set in Signature Challenge Training Exercises

Set #	Vehicles in Set	Identification Score (% correct)		Response Time (sec)	
		Near Practice	Far Practice	Near Practice	Far Practice
1	M1	93	82	2.20	2.63
	T72	92	79	2.20	2.95
2	M2	99	90	1.50	2.19
	BMP	92	75	1.81	2.38
3	LAV	87	76	2.18	2.65
	BTR	87	75	2.07	2.63
4	M1	94	87	2.04	1.99
	T72	89	79	2.17	2.33
5	M2	99	88	1.41	1.98
	BMP	88	63	1.76	2.31
6	LAV	91	82	1.75	2.14
	BTR	89	82	1.82	2.11
7	M1	96	87	1.48	1.37
	M2	98	91	1.24	1.34
	LAV	89	89	1.55	1.75
8	T72	93	82	1.59	1.83
	BMP	85	83	2.09	2.24
	BTR	87	72	1.71	1.87
All Sets		92	81	1.81	2.15

Note. For overall mean, n is 264 trials for near practice; 272 trials for far practice.

Table D-7

Experiment 3: Mean Identification Scores (% correct) for Front and Rear Aspects (no transfer) on the Training Exercises and Tests

Aspect	Vehicle	**Training Exercise -** Range Band		**Test -** Range Band	
		Near	Far	Near	Far
		Vehicle ID			
Front	M1	96	58	94	74
	T72	83	83	88	79
	M2	90	69	94	79
	BMP	55	27	73	44
	LAV	69	52	85	68
	BTR	45	31	54	47
Rear	M1	76	72	76	94
	T72	69	66	100	85
	M2	83	55	91	74
	BMP	59	38	64	65
	LAV	24	31	39	59
	BTR	82	38	97	71
		Signature Challenge			
Front	M1	91	41	88	30
	T72	81	53	73	61
	M2	97	88	94	76
	BMP	72	74	61	42
	LAV	97	97	76	91
	BTR	64	26	67	45
Rear	M1	100	97	91	91
	T72	88	79	94	88
	M2	97	85	94	64
	BMP	76	88	61	64
	LAV	27	85	27	36
	BTR	70	41	85	76

Note. There was only one session of the Vehicle ID Training Exercise; all 6 vehicles were presented as choices. Signature Challenge Training Exercise percentages reflect the 7th and 8th sets (US and nonUS, respectively). Signature Challenge percentages are based on each soldier's first session in each set. In both the Vehicle ID and Signature Challenge tests, all six vehicles were presented as choices. For the front and rear aspects, soldiers were trained and tested at the same range band. Percentages based on 33 soldiers for Near range band; 34 for Far range band.

Table D-8

Experiment 3. Mean Identification Scores for Significant Effects on Vehicle ID and Signature Challenge Tests: Front and Rear Aspects (no transfer)

Effect	Mean % Correct					
Practice[a]	Near			Far		
	77			66		
Test[b]	Vehicle ID			Signature Challenge		
	75			69		
Vehicle[c]	M1	M2	T72	BMP	BTR	LAV
	79	83	83	59	67	60
Practice x Vehicle[d]	M1	M2	T72	BMP	BTR	LAV
- Near	87	93	89	64	76	57
- Far	72	73	78	54	59	62
Practice x Instruct x Vehicle[e]	M1	M2	T72	BMP	BTR	LAV
Near Pract & Near Instr	87	94	87	69	81	53
Near Pract & Far Instr	87	92	91	60	70	61
Far Pract & Near Instr	74	78	76	46	61	79
Far Pract & Far Instr	70	69	80	63	58	45
Practice x Test x Vehicle[f]	M1	M2	T72	BMP	BTR	LAV
Near Pract w Veh ID	85	92	94	68	76	62
Near Pract w SC	89	94	83	60	75	51
Far Pract w Veh ID	84	76	83	55	59	62
Far Pract w SC	60	70	73	53	60	62
Practice x Instruct x Test x Vehicle[g]	M1	M2	T72	BMP	BTR	LAV
Veh ID						
Near Pract & Near Instr	79	97	91	76	73	56
Near Pract & Far Instr	91	87	97	59	78	69
Far Pract & Near Instr	86	75	75	44	61	80
Far Pract & Far Instr	81	78	91	66	56	44
SC						
Near Pract & Near Instr	94	91	82	62	88	50
Near Pract & Far Instr	84	97	84	59	62	53
Far Pract & Near Instr	61	80	78	47	61	78
Far Pract & Far Instr	59	59	69	59	59	47

Note. $N = 67$ soldiers. Each soldier was trained & tested at the same range band.

[a] $F(1, 63) = 12.92, p < .0006$

[b] $F(1, 63) = 9.48, p < .0030$

[c] $F(5, 315) = 15.90, p < .0000$

[d] $F(5, 315) = 2.38, p < .0382$

[e] $F(5, 315) = 3.83, p < .0022$

[f] $F(5, 315) = 2.37, p < .0389$

[g] $F(5, 315) = 2.87, p < .0165$

Table D-9

Experiment 3: Mean Identification Scores (% correct) on the Transfer Imagery (oblique aspects) from the Training Exercises and Tests as a Function of Vehicle and Aspect

Vehicle	Mean % Correct											
	SC Training				SC Test				Vehicle ID Test			
	Near Practice Condition (*n* = 33 soldiers per vehicle per aspect)											
	Near Images (no transfer)				Far Images (transfer)				Far Images (transfer)			
	Aspect Presented				Aspect Presented				Aspect Presented			
	Lf Ft	Rt Ft	Lf R	Rt R	Lf Ft	Rt Ft	Lf R	Rt R	Lf Ft	Rt Ft	Lf R	Rt R
M1	88	100	97	100	61	88	88	97	61	73	94	85
T72	100	88	91	94	62	42	67	70	79	61	88	73
M2	100	97	97	100	82	97	91	94	79	100	85	70
BMP	94	91	88	88	67	40	25	55	40	85	70	52
BTR	100	85	94	97	91	33	75	84	81	70	94	49
LAV	94	97	100	100	50	55	64	79	45	91	82	64
	Far Practice Condition (*n* = 34 soldiers per vehicle per aspect)											
	Far Images (no transfer)				Near Images (transfer)				Near Images (transfer)			
	Aspect Presented				Aspect Presented				Aspect Presented			
	Lf Ft	Rt Ft	Lf R	Rt R	Lf Ft	Rt Ft	Lf R	Rt R	Lf Ft	Rt Ft	Lf R	Rt F
M1	74	97	94	94	91	88	91	82	82	85	85	88
T72	82	82	76	91	82	54	82	94	94	91	82	88
M2	91	94	91	82	88	91	94	97	91	97	97	94
BMP	85	76	94	70	85	70	87	61	85	85	76	59
BTR	97	57	91	65	73	88	88	85	85	88	91	85
LAV	85	76	91	82	70	97	94	88	85	85	94	91

Note. Signature Challenge Training scores are from the first session of Vehicle Sets 7 and 8. There was no Vehicle ID training on the oblique aspects.

Table D-10

Experiment 3. Mean Identification Scores for Significant Effects on Vehicle ID and Signature Challenge Tests: Oblique Aspects (transfer)

Effect	**Mean % Correct**					
Practice[a]	Far (Near image tested)			Near (Far image tested)		
	85			71		
Test[b]	Vehicle ID			Signature Challenge		
	80			76		
Vehicle[c]	M1	M2	T72	BMP	BTR	LAV
	84	89	76	64	78	76
Practice x Vehicle[d]	M1	M2	T72	BMP	BTR	LAV
- Far (Near image tested)	86	92	84	76	85	87
- Near (Far image tested)	81	87	68	53	72	65
Test x Vehicle[e]	M1	M2	T72	BMP	BTR	LAV
- Veh ID	82	88	83	69	81	79
- SC	85	91	69	59	76	73
Instruct x Practice x Test[f]	Near Instruct Far Practice (Near image tested)		Near Instruct Near Practice (Far image tested)		Far Instruct Near Practice (Far image tested)	Far Instruct Far Practice (Near image tested)
Veh ID	89		74		71	86
SC	88		65		72	77

Note. $N = 67$

[a] $F(1, 63) = 22.21, p < .0000$

[b] $F(1, 63) = 9.52, p < .0030$

[c] $F(5, 315) = 15.02, p < .0000$

[d] $F(5, 315) = 3.09, p < .0096$

[e] $F(5, 315) = 5.14, p < .0001$

[f] $F(1, 63) = 8.63, p < .0046$

Table D-11

Experiment 3: Aspect Confusion Matrix for Vehicle ID Test

Aspect Response & Practice Condition	Aspect Presented and Range Band of Images Displayed in Vehicle ID Test					
	Near Images (no transfer)		Far Images (transfer)			
Near Practice	Front	Rear	Lf Front	Lf Rear	Rt Rear	Rt Front
Front	**88**	11	3	1	5	0
Rear	10	**89**	2	1	12	0
Lf Front	3	0	**33**	5	8	2
Lf Rear	0	0	8	**79**	4	5
Rt Rear	0	0	6	1	**47**	9
Rt Front	0	0	6	4	3	**53**
Lf Flank	0	0	40	9	9	1
Rt Flank	0	0	4	2	12	31
	Far Images (no transfer)		Near Images (transfer)			
Far Practice	Front	Rear	Lf Front	Lf Rear	Rt Rear	Rt Front
Front	**65**	24	1	0	0	0
Rear	27	**71**	1	1	5	0
Lf Front	1	0	**75**	0	6	0
Lf Rear	1	1	4	**92**	0	6
Rt Rear	1	4	4	0	**84**	0
Rt Front	1	0	1	3	0	**85**
Lf Flank	1	0	11	3	0	1
Rt Flank	1	0	2	0	3	7

Note. Percent correct responses for aspects is in bold. Columns may not sum to 100 because of rounding. Some soldiers indicated flanks were displayed, even through there were no flanks presented. Percentages for each aspect presented in the Near Practice condition were based on 198 trials (33 soldiers, 6 vehicles). Percentages in the Far Practice condition were based on 204 trials (34 soldiers, 6 vehicles).

Table D-12
Experiment 3: Mean Identification Scores for Significant Instruction and Practice Effects on Pretests and Posttests

Factor	**Mean % Correct**					
	Visible		Difference:	FLIR		Difference:
	Pre	Post	Post-Pre	Pre	Post	Post-Pre
Practice x Pre-post x Vis-FLIR[a]						
Near	69	80	11	57	81	24
Far	74	81	7	64	78	14
Practice x Pre -post x Train[b]	No Train			Train		
	Pre	Post		Pre	Post	
Near	66	72	6	60	88	28
Far	71	76	5	68	84	16
Instruct x Practice x Vis-FLIR x Pre-post[c]	Visible			FLIR		
	Pre	Post		Pre	Post	
Instruct -Near Practice						
Near	66	76	10	52	80	28
Far	75	81	6	65	77	12
Instruct - Far Practice						
Near	72	84	12	62	81	19
Far	74	82	8	63	79	16

Note. $N = 67$. Practice x Pre-post $F(1, 63) = 7.26, p < .0090$
[a] $F(1, 63) = 4.48, p < .0382$
[b] $F(1, 63) = 7.35, p < .0086$
[c] $F(1, 63) = 7.54, p < .0078$

For the 3-way interaction with practice, pre-post and train-no train (see Table D-12), the initially higher scores on visible imagery for those soldiers in the far practice group (vs near practice) was not maintained. However, for the FLIR imagery, those in the near practice group increased more (24%) than those in the far group (14%). The near group had the higher posttest scores; the far group had the higher pretest scores. This result parallels the other interaction involving the practice factor, and may reflect a simple transfer effect from practicing on imagery presented at the same range during the posttest.

Table D-13

Experiment 3: Vehicle Confusion Matrix from Pretests and Posttests

Vehicle Response	Vehicle Presented									
	M1	T72	M60	M2	BMP	LAV	BTR	M113	ZSU	HMMWV
M1 PreVis	**93**	8	2	4	4	0	0	1	7	0
PreIR	**95**	9	2	1	2	0	0	0	10	0
PostVis	**77**	10	2	6	4	2	0	1	10	0
PostIR	**90**	4	1	1	1	1	0	0	0	0
T72 PreVis	1	**67**	9	0	4	0	1	0	4	0
PreIR	2	**83**	9	1	5	1	1	0	4	0
PostVis	7	**65**	13	0	4	1	1	0	6	0
PostIR	4	**90**	10	0	5	0	1	0	13	0
M60 PreVis	2	7	**65**	0	4	0	0	0	8	0
PreIR	0	2	**67**	0	1	0	1	0	5	0
PostVis	5	9	**51**	0	4	1	0	0	16	1
PostIR	0	2	**60**	0	1	1	1	0	9	0
M2 PreVis	1	0	1	**93**	4	5	1	0	6	0
PreIR	2	0	0	**97**	1	4	1	0	4	0
PostVis	5	2	2	**83**	23	24	1	1	13	0
PostIR	3	0	1	**94**	6	16	0	0	8	0
BMP PreVis	0	3	1	0	**62**	7	13	1	3	0
PreIR	0	2	2	0	**78**	1	11	0	9	0
PostVis	0	3	1	0	**48**	7	12	2	5	0
PostIR	1	1	1	3	**84**	1	2	0	7	0
LAV PreVis	0	0	1	1	2	**61**	23	0	2	0
PreIR	0	0	1	1	0	**85**	10	0	1	0
PostVis	0	0	2	1	2	**36**	34	1	1	1
PostIR	0	0	1	0	0	**73**	5	0	1	0
BTR PreVis	0	2	2	0	5	17	**48**	0	6	0
PreIR	1	1	0	0	2	7	**73**	0	0	0
PostVis	1	2	3	1	5	22	**40**	1	8	0
PostIR	0	1	2	0	1	8	**91**	1	1	0
M113 PreVis	0	0	1	0	0	1	0	**98**	1	0
PreIR	0	0	1	0	1	0	0	**99**	1	0
PostVis	0	0	0	1	1	1	2	**96**	2	0
PostIR	0	1	1	0	0	0	1	**98**	1	0
ZSU PreVis	0	1	1	0	4	3	5	0	**34**	0
PreIR	0	1	2	0	1	1	2	1	**43**	0
PostVis	1	2	4	4	2	3	2	0	**18**	0
PostIR	0	1	2	1	0	1	1	1	**34**	0
HMMWV										
PreVis	0	0	0	0	0	0	1	0	0	**96**
PreIR	0	0	0	0	0	0	0	0	0	**96**
PostVis	0	0	1	0	0	2	1	0	0	**90**
PostIR	0	0	0	0	0	0	0	0	0	**94**
M551 PreVis	1	8	14	0	8	1	1	0	9	0
PreIR	0	2	13	0	2	0	0	0	10	0
PostVis	3	5	17	1	5	0	2	0	9	0
PostIR	0	1	17	0	1	0	0	0	9	0
Other[a] PreVis	1	2	3	1	2	4	8	0	18	4
PreIR	0	0	2	0	0	0	1	0	12	4
PostVis	1	1	1	3	1	1	4	1	10	8
PostIR	1	0	2	0	0	0	0	0	12	6

In Table D-13, the diagonal cells contain the percent correct responses. The off-diagonal cells represent the percentage of instances where the vehicle presented was confused with, or misidentified as, another vehicle. The columns sum to 100% within rounding error. In the pretests and posttests, each vehicle was displayed 3 times; each presentation was a different aspect angle. The category of "other refers to the M35 2 ton truck, M814 5 ton truck, and the ZIL-131 truck. Number of trials per vehicle was 201 (67 soldiers, 3 displays per vehicle).

Table D-14

Experiment 3: Response Times (sec) for Significant Effects on the Pretest and Posttest

Effect	Mean Time *M* (*SD*)	
Imagery[a]	Visible	FLIR
	6.53 (1.43)	11.65 (2.39)
Pre-Post[b]	Pretest	Posttest
	11.13 (2.25)	7.04 (2.09)
Imagery x Pre-post[c]	Pretest	Posttest
Visible	8.01 (2.77)	6.07 (1.79)
FLIR	15.28 (3.91)	6.99 (2.78)

$N = 67$.

[a] $F(1, 63) = 150.41, p < .0000$

[b] $F(1, 63) = 438.71, p < .0000$

[c] $F(1, 63) = 108.95, p < .0000$

Table D-15
Experiment 3: Pretest and Posttest Correlations

	Scores				Response Time			
	Visible		FLIR		Visible		FLIR	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Scores								
VisPre	--	.83***	.86***	.70***	-.37**	-.48***	-.21	-.15
VisPost		--	.70***	.85***	-.30*	-.38**	-.00	-.23
FLIRPre			--	.55***	-.39**	-.38**	-.16	-.08
FLIRPost				--	-.26*	-.35**	.01	-.28*
Time								
VisPre					--	.44***	.19	.29*
VisPost						--	.29*	.66***
FLIRPre							--	-.00
FLIRPost								--

Note. $N = 67$.
*** $p < .001$. ** $p < .01$. * $p < .05$.

Table D-16
Experiment 3: Correlations Among Scores on the Training Exercises and Tests

		Training Exercises		Tests: Transfer		Tests: No Transfer	
	E&E Test	Veh ID	SC: # Sessions	Veh ID	SC	Veh ID	SC
E&E	--	.33**	-.09	.32**	.13	.27*	.30*
Train							
ID		--	-.58***	.36**	.22	.66***	.61***
SC: # Sess			--	-.09	.08	-.50***	-.50***
Tests: Transfer							
ID				--	.67***	.32**	.22
SC					--	.18	.11
Tests: No Trans							
ID						--	.58***
SC							--

Note. $N = 67$.
*** $p < .001$. ** $p < .01$. * $p < .05$.

Table D-17

Experiment 3: Correlations Between the Pre- and Posttests and Scores on the Training Exercises and Tests

	Training and Test Scores						
		Training Exercises		Tests: Transfer		Tests: No Transfer	
Pre- and Posttests	E&E Test	Veh ID	SC: # Sessions	Veh ID	SC	Veh ID	SC
Scores							
VisPre	.35**	.43***	-.22	.66***	.58***	.37**	.30*
VisPost	.38**	.52***	-.30*	.59***	.57***	.48***	.38**
FLIRPre	.39***	.32**	-.21	.62***	.51***	.36**	.27*
FLIRPost	.32**	.65***	-.43***	.59***	.46***	.51***	.43***
Time							
VisPre	-.26*	-.12	.08	-.41***	-.26*	-.19	-.27*
VisPost	-.25*	-.13	-.05	-.24*	-.32**	-.15	-.08
FLIRPre	-.15	.05	-.14	-.23	-.04	.12	.08
FLIRPost	-.09	-.07	-.13	-.12	-.31	-.09	-.07

Note. $N = 67$.

*** $p < .001$. ** $p < .01$. * $p < .05$.

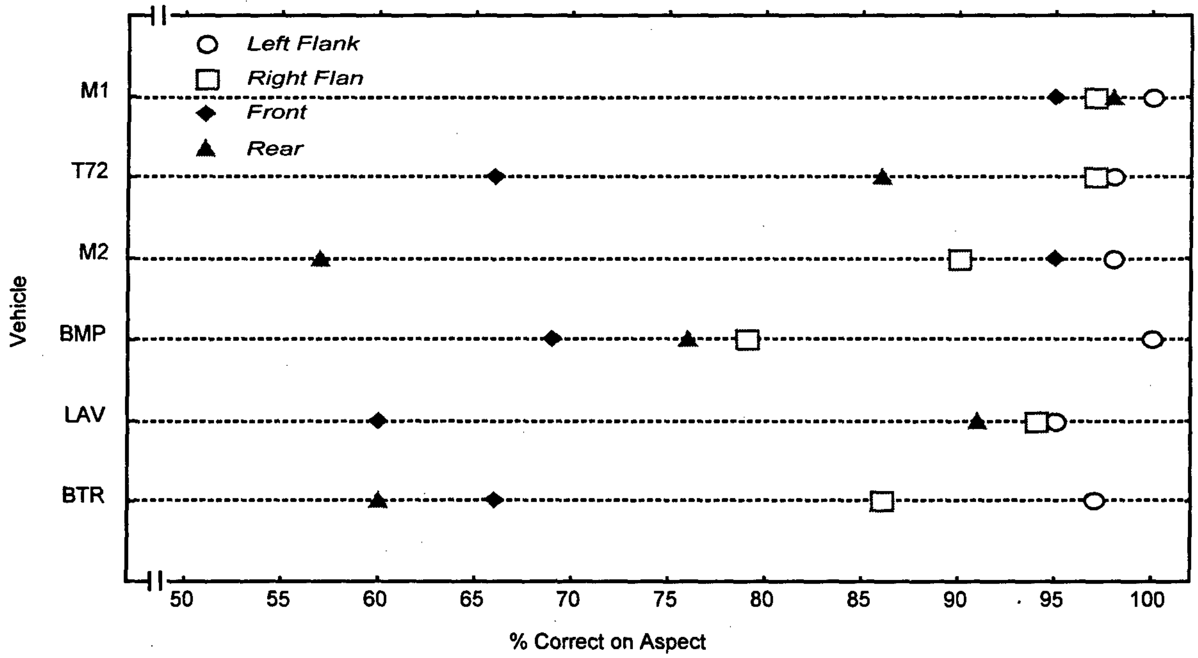

Figure D-1. Experiment 3: Aspect scores on Vehicle ID training exercises as a function of vehicle and aspect presented.

Discussion of Figures D-2, D-3, and D-4

Transfer Data:

Significant Interactions on Identification Scores from the Repeated Measures ANOVA

A vehicle by practice interaction showed that the difference between the near and far imagery scores was smallest for the M1 and M2 (5% points difference) and greatest for the BMP and LAV (25% points difference, Figure D-2). Vehicle scores also varied with the test (Figure D-3). Vehicle ID scores were about 12% points higher than Signature Challenge for the T72 and BMP, and 5% points higher for the BTR and LAV; but were 3% points lower for the M1 and the M2.

Finally, although Vehicle ID scores were generally higher than Signature Challenge scores, this difference was affected by the initial instruction and the follow-on training exercises. Vehicle ID scores were higher than Signature Challenge when the range band of the instruction and practice imagery differed from the test range band (Figure D-4). But there was practically no difference between the two tests when the initial instruction corresponded to the range band in the test. The advantage of the Vehicle ID test format in this case may have occurred because Vehicle ID gave soldiers time to study the imagery as well as the option to change their response. Both factors could have contributed to higher scores when the soldier had never been exposed to imagery at the transfer range prior to the transfer test.

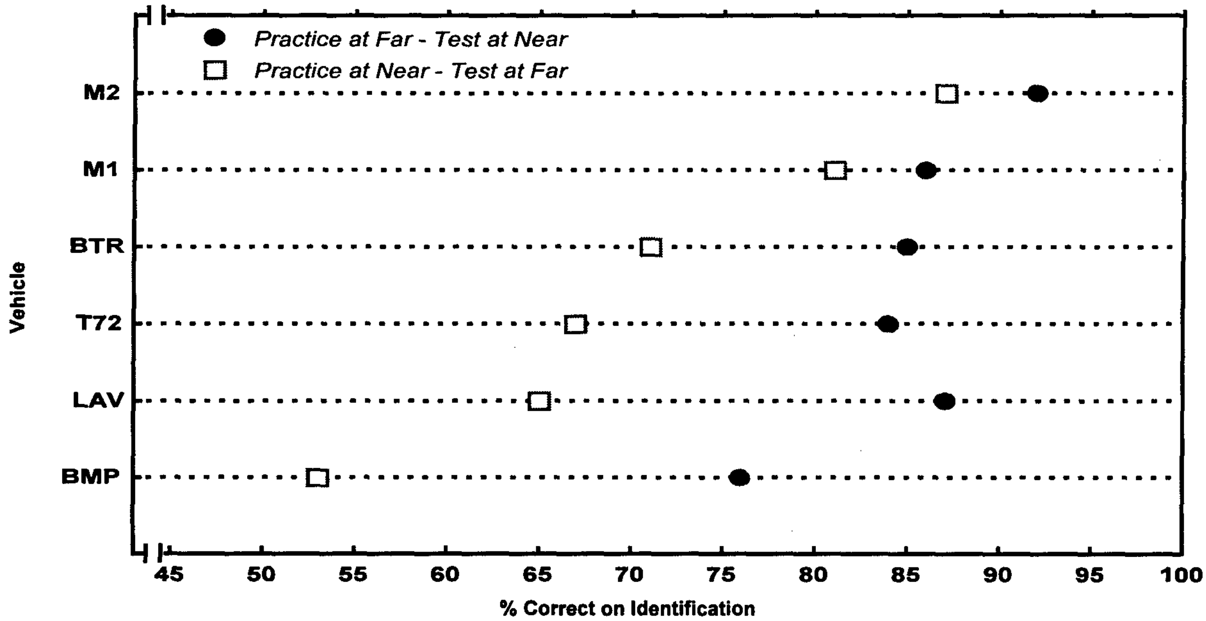

Figure D-2. Experiment 3: Identification scores on tests for oblique aspects (transfer) as a function of vehicle and practice-to test range band.

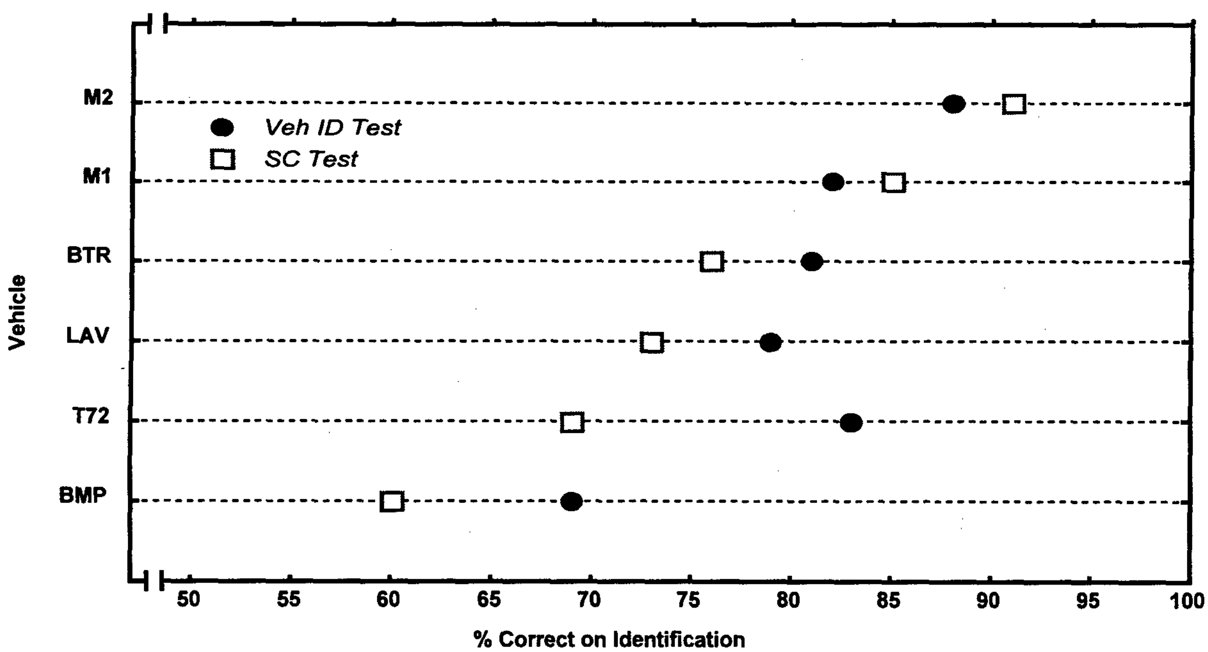

Figure D-3. Experiment 3: Identification scores on tests for oblique aspects (transfer) as a function of vehicle and type of test.

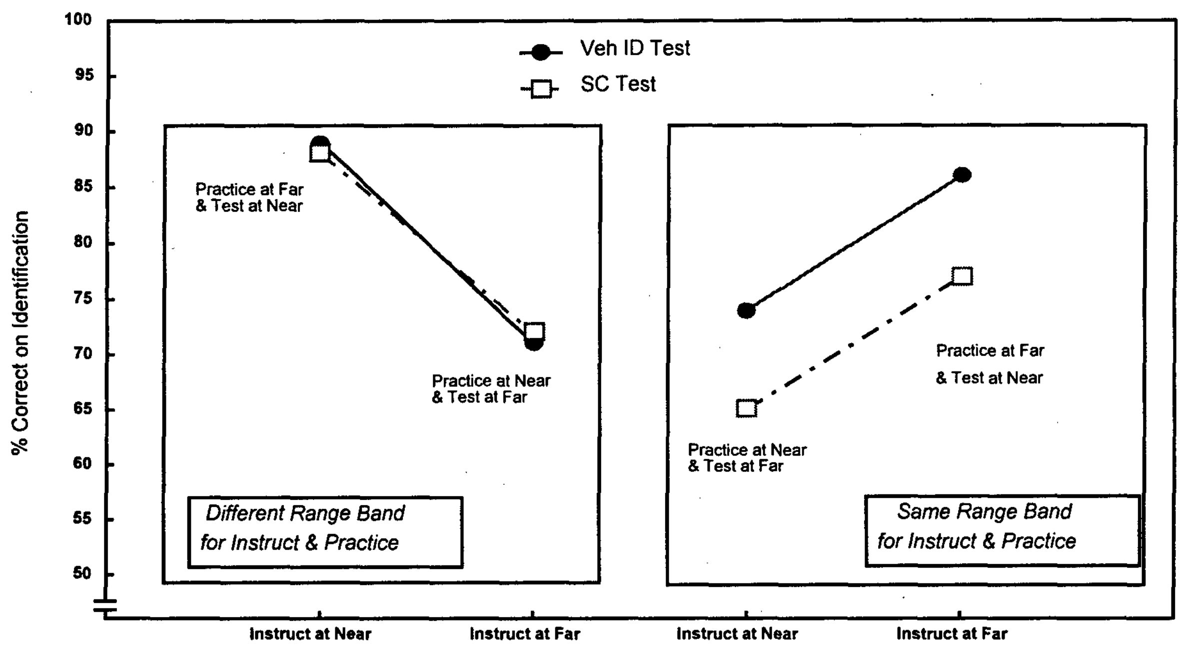

Figure D-4. Experiment 3: Differences in Vehicle ID and SC test scores on oblique aspects as a function of consistency of image range during instruction and practice.

APPENDIX E

SOLDIER REACTIONS TO CVI*PLUS*

Did you like the CVI*Plus* Program?

Yes - 100% (66/66)

Are you better at identifying vehicles with thermal sights after CVI*Plus* training than before you had this training?

Yes - 100% (66/66)

Would you use the program if you had it in your unit?

Yes - 100% (66/66)

Which part did you like the best? (soldiers frequently checked more than one module)

Image Library -	24.24% (16/66)
Signature Challenge	77.27% (51/66)
Vehicle ID -	42.42% (28/66)

One comment - "All are excellent training aids."

What changes would you recommend?

No Recommendations - 21 comments

None- 14

None, except I would like to spend more time training on this program. I have to improve my vehicle identification.

None, just make this system more available for the soldier. It is a great tool to train.

I wouldn't make any changes because this is the first time I have done this and I think it is the best way to ID targets.

With this being my first time using this system, I benefited very much. It's not everyday my unit get to train with such a system. I'm sure the system will improve with time, but as of now I'm satisfied with your system.

There's not really any changes that I can see that really could be made.

I don't see anything that I would really change about the system. It does a good job of showing how to pick features of different vehicles.

No changes [maybe add more time to class].

Image Quality - 5 comments

Make image sharper.

Make them clearer.

If possible, a clearer day time image.

Better graphics. The images are a little fuzzy.

Change the night sight. Can't put my finger on it, but there is something funny with it.

BH and WH Flexibility -2 comments

Put BH and WH interchangeable in the program.

Choice between white hot and black hot.

Tests and Scoring - 5 comments

5 minute breaks between tests. Test scores sent to units.

Show score after scored test. Make an interactive timed test with BFV controls where you must choose friend or foe.

Be able to know the test results after test.

On Signature Challenge, you should have 100% correct before moving on. In war the 1% that you get wrong might be a friendly vehicle and that is unacceptable.

Ability to pause during timed test.

Vehicle Distance/Range - 2 comments

The long ranges are not necessary for two reasons: (1) As 11M gunners, we are trained to not engage targets until we can identify vehicles positively, (2) Another point is long ranges are past the Bradley's maximum effect ranges of all weapon systems, or at least they appear to be.

Distance, not range 1-2-3.

Program Design - 3 comments

Better menu selection, i.e., drop-down menus.

Showing visible to thermal relation as part of the instruction would make identification easier. (*With this group, the instructor did not include side-by-side comparisons of visible and thermal, as in the previous experiments*).

Instead of a LAV-25 thermal basics image only, a BFV thermal image too.

Vehicles -13 comments

Add some more vehicles - 7

More threat vehicles such as T62, T-54/55, BRDM, and MTLB.

Vehicles that are in the inventory of countries like North Korea, Iraq, and Iran.

That you use vehicles that are rarely seen and use vehicles with similar signatures.

Moving vehicles and land features.

More time and vehicles to train on.

A battle simulation with all the vehicles.

Other Comments

Work on the glitches.

Headphones for audio.

Add target sightings.

Magnification option.

Get the crews together for classes.

More classroom time, then go to hands-on.

Do this more often.

More training on the program.

Signature Challenge.

What other features would you like included?

No recommendations - 10 comments

Nothing- 8

None, don't know enough about the program. Seems pretty good as is.

Everything is fine the way it is.

Sight Features - 9 comments

Range finder - to give the range through the ISU. -2.

A focus knob -2

The only feature I can think of is to vary between high magnification and low magnification to improve this system for Bradley gunners.

BFV sight training.

Perhaps also a test that uses sights from the M1A2 or the M2A3.

An estimate range scenario using vehicle images with the Bradley choke site reticle overlaid on top of vehicle.

Sighting and tracking.

More Information on Vehicles - 3 comments

Characteristics of vehicles, e.g., weapon systems, engine, etc.

Weapon systems, size ammo, more information about the vehicles.

A library of all combat vehicles and their signatures, an encyclopedia of sorts - text and schematic.

More Vehicles, Aircraft, Moving Vehicles, Battlefield Realism - 16 comments

There should be more vehicles on the system.

More enemy and friendly vehicles.

More vehicles; add aircraft.

Moving vehicles, and air targets, i.e., Apache vs. Hind, ground troops and equipment - US vs. Soviet.

More vehicles, vehicles dug into fighting positions.

Include obscured vehicles in the imagery.

Vehicles in defilade positions with just the turret exposed. Also identifying moving vehicles should be included.

Moving vehicles much like in the demo. Some aircraft.

Moving vehicles would make it more realistic.

Video images, more images, battle realism.

Moving targets; battlefield conditions.

Enemy air (Hind); friendly air.

Add air vehicles, helicopters, jet planes. I would like to say this is a very good program.

Helicopters and airplanes.

Maybe a battlefield scenario with multiple enemy and friendly targets included in your ID challenge.

I think maybe adding an engagement practice, so the student can practice the correct threat vehicle from friendly.

Other Comments

I did like the class, but we need more training/ constant training on vehicle ID.

Not sure.

Maybe two different sides of vehicle to get a better view.

Personally, I would recommend this system to not only the mechanized units, but light infantry as well. If this system was available to light infantry (Dragon and Javelin gunners), then I would think that we would have a more solid understanding that we can give the foot soldier the knowledge to identify US and enemy vehicles. The program itself is very solid and it does give out a lot of knowledge. Two thumbs up!

Whatever new stuff you get.

Increasing the depth in the picture.

APPENDIX F

CVI*Plus* MODULES

Figures

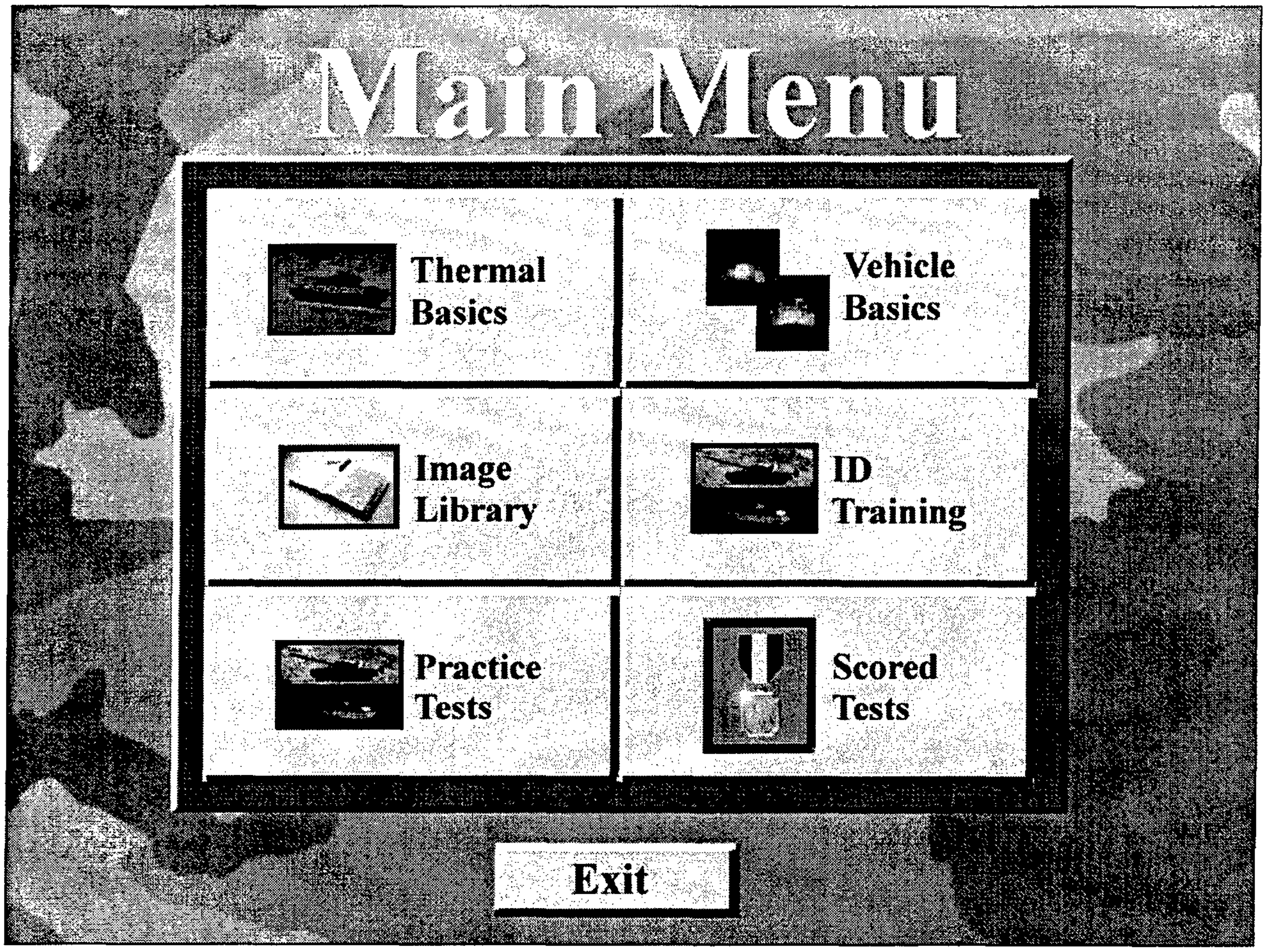

Figure F-1. Main Menu. This display represents the Main Menu portion of CVI*Plus* including buttons for Thermal Basics, Vehicle Basics, Image Library, ID Training, Practice Tests (Exercises), and Scored Tests.

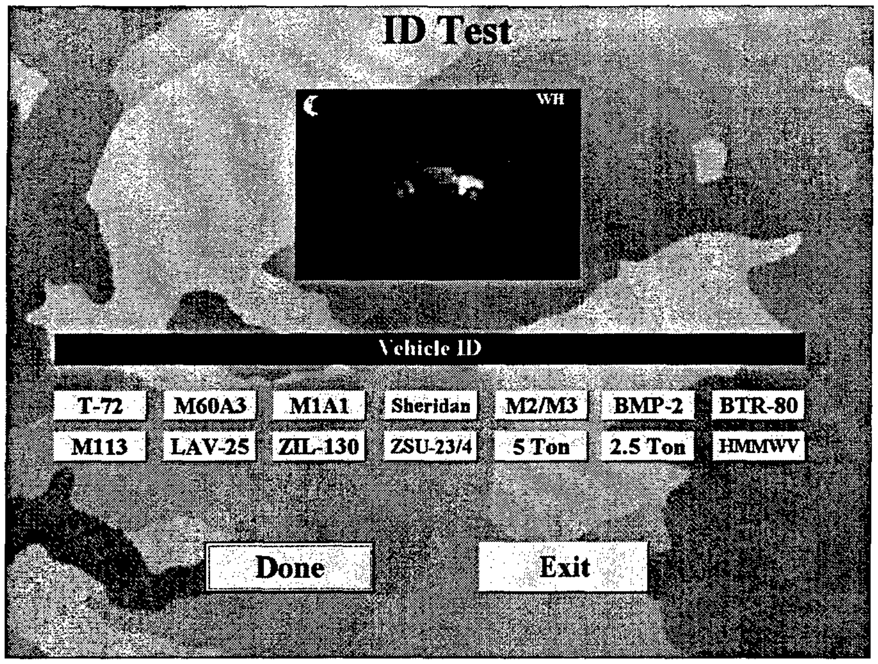

Figure F-2. Pretest/Posttest Display, Thermal (FLIR). This display represents the thermal signature of a HMMWV from the right flank, in white hot mode at night.

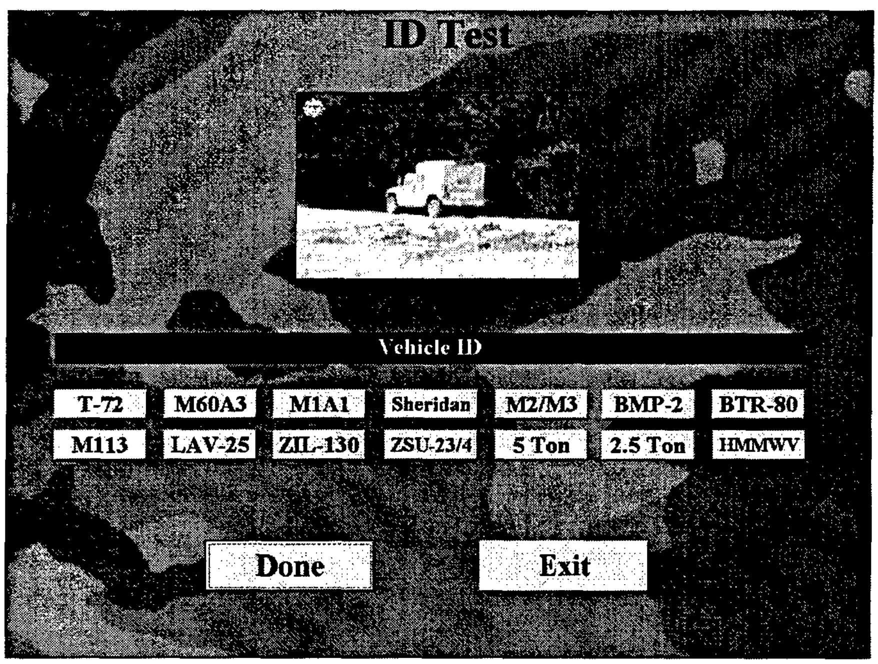

Figure F-3. Pretest/Posttest Display, Visible. This display contains a picture of a HMMWV from the left rear oblique.

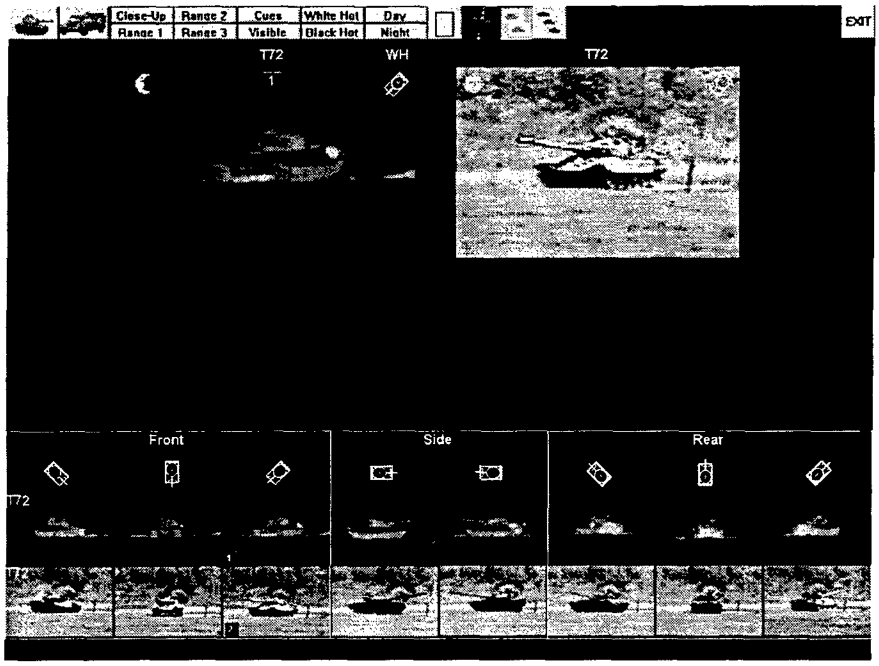

Figure F-4. Image Library. This display compares the thermal signature of a T72 (white-hot, night) to a visual picture of a T72. Both images are of the left front oblique.

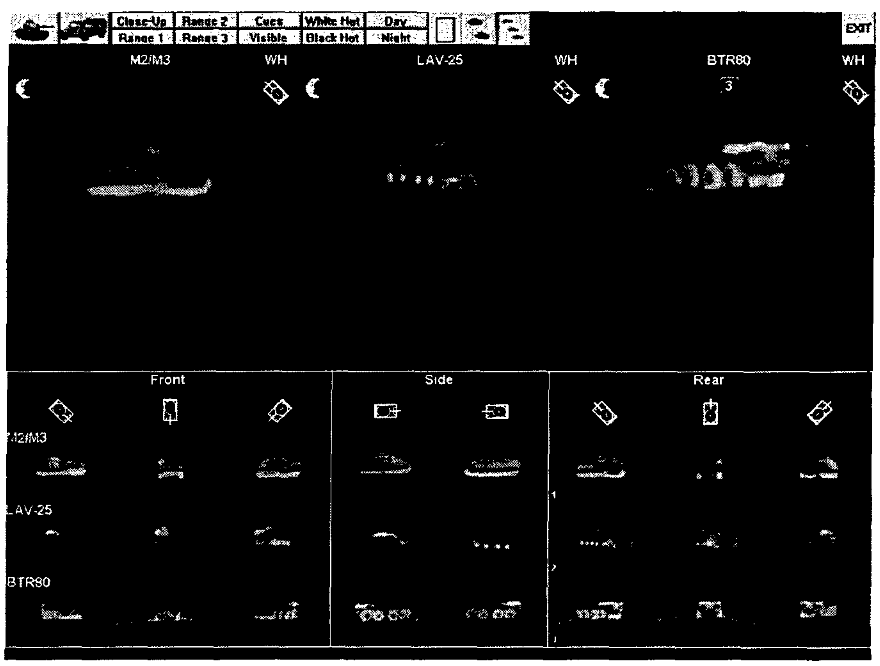

Figure F-5. Image Library: This display compares the thermal signatures of an M2/M3, a LAV-25, and a BTR-80. All represent the left rear oblique at night using white-hot FLIR.

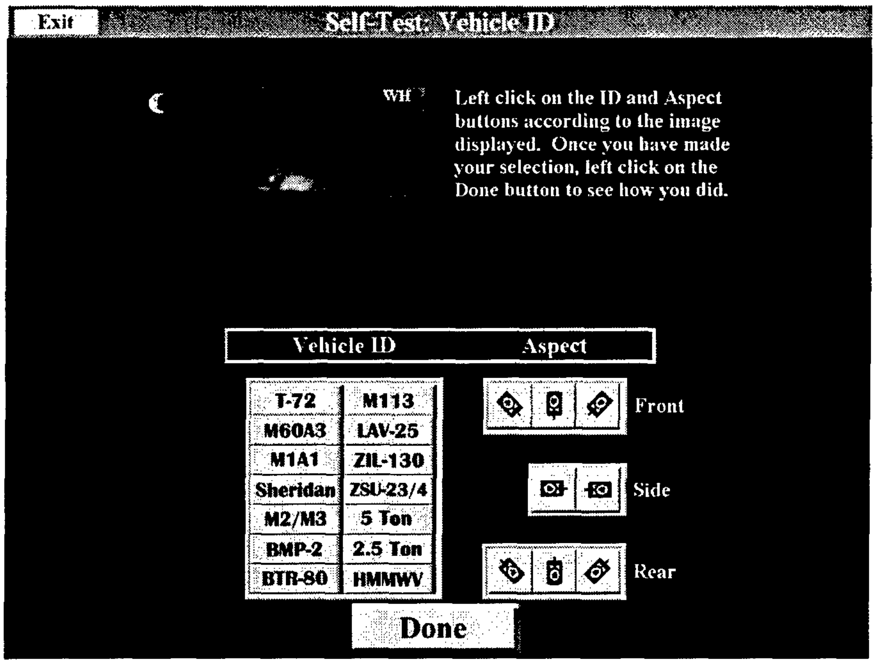

Figure F-6. Vehicle ID: This display represents the Vehicle ID module before a selection is made. Image shown is a T72 (white-hot, night) from the left rear oblique.

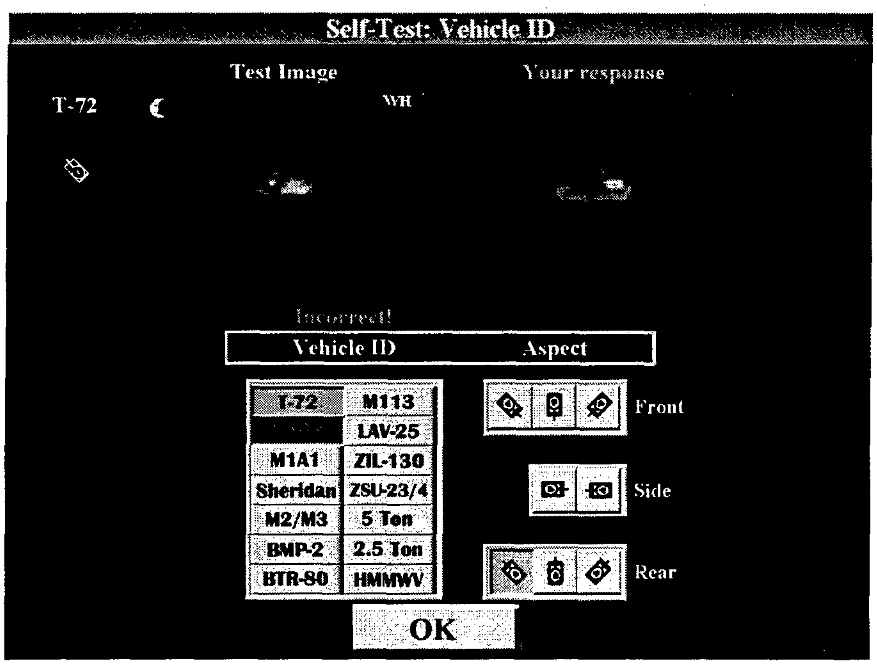

Figure F-7. Vehicle ID: This display represents the Vehicle ID module after an incorrect selection has been made. Images shown represent the T72 (test image) and the M60A3 (soldier's response) left rear oblique (white-hot, night).

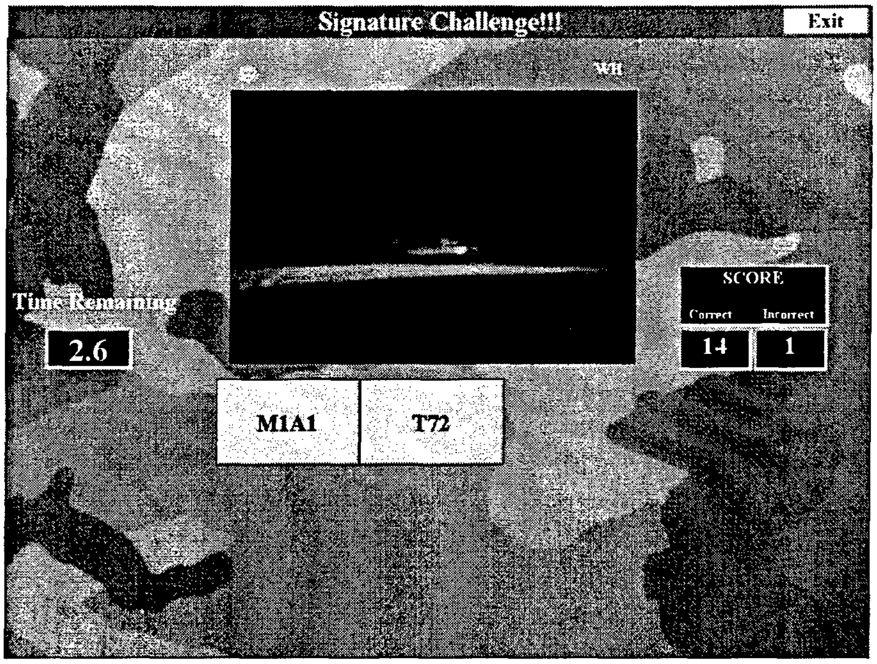

Figure F-8. Signature Challenge: This image represents the Signature Challenge display, in an exercise training discrimination of two vehicles, before a selection is made. Image shown is the left flank of a T72 (white-hot, day).

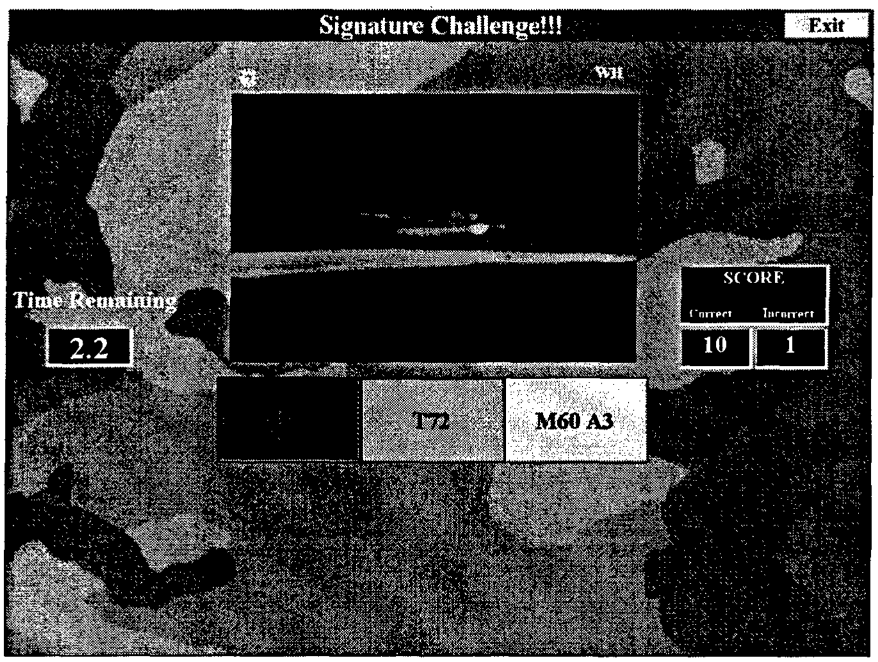

Figure F-9. Signature Challenge: This image represents the Signature Challenge display, in an exercise training discrimination of three vehicles, after an incorrect selection (M1A1) is made. Buttons chosen incorrectly turn red. Image shown is a T72 left flank (white hot, day).